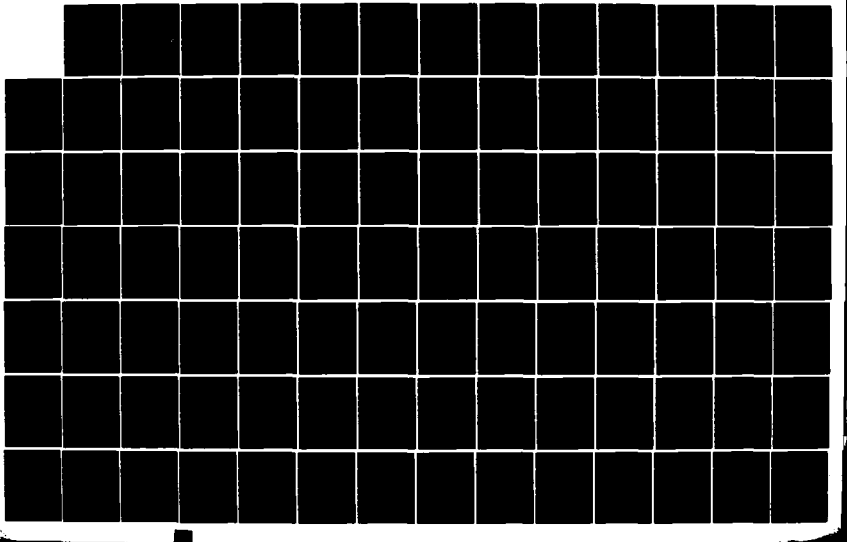
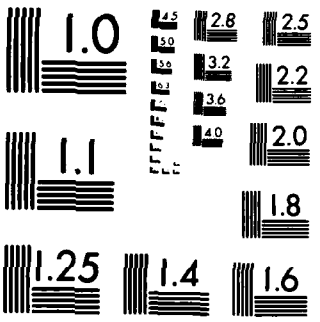
AD-A152 070 DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA(U) BATTELLE COLUMBUS LABS OH P I FEDER DEC 81 BATT-G-7673 DAMD17-80-C-0165 1/5

UNCLASSIFIED F/G 6/20 NL





PHOTOGRAPH THIS SHEET

AD-A152 070

DTIC ACCESSION NUMBER

LEVEL

1

INVENTORY

Final Report - Dec 1981

DOCUMENT IDENTIFICATION DAMD17-80-C-0165

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DISTRIBUTION STATEMENT

ACCESSION FOR		
NTIS	GRA&I	X
DTIC	TAB	☐
UNANNOUNCED		☐
JUSTIFICATION		
BY		
DISTRIBUTION /		
AVAILABILITY CODES		
DIST	AVAIL AND/OR SPECIAL	
A-1		

DISTRIBUTION STAMP

DTIC
COPY
INSPECTED

DTIC FILE COPY

DTIC
ELECTE
S MAR 29 1985 D
D

DATE ACCESSIONED

DATE RETURNED

DATE RECEIVED IN DTIC

REGISTERED OR CERTIFIED NO.

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-DDAC

DTIC FORM 70A DEC 83

DOCUMENT PROCESSING SHEET

PREVIOUS EDITION MAY BE USED UNTIL STOCK IS EXHAUSTED.

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AD _____

DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH *DAPHNIA MAGNA*

Final Report

by

Paul I. Feder
Applied Statistics Group, Battelle Columbus Laboratories

December, 1981

Supported by

U.S. Army Medical Research and Development Command
Fort Detrick, Frederick, Maryland 21701

Contract No. DAMD17-80-C-0165

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

Contract Officer's Technical Representative: Paul H. Gibbs
U.S. Army Medical Bioengineering Research and Development Laboratory
Fort Detrick, Frederick, Maryland 21701

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA		5. TYPE OF REPORT & PERIOD COVERED FINAL REPORT Nov. 1, 1980 - Dec. 31, 1981 6. PERFORMING ORG. REPORT NUMBER G-7673
7. AUTHOR*(s)* Paul I. Feder		8. CONTRACT OR GRANT NUMBER*(s)* DAMD17-80-C-0165
9. PERFORMING ORGANIZATION NAME AND ADDRESS BATTELLE, Columbus Laboratories 505 King Avenue Columbus, Ohio 43201		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102A.3E161102BS04.AA.050
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Medical Research and Development Command Fort Detrick, Frederick, Maryland 21701		12. REPORT DATE December 31, 1981 13. NUMBER OF PAGES 464
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* U.S Army Medical Bioengineering Research and Development Laboratory Fort Detrick, Frederick, Maryland 21701		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Daphnia Magna	confidence intervals	outlier detection
chronic aquatic toxicity tests	dose response estimation	variance
statistical analysis	multiple regression analysis	heterogeneity
graphical data display	computer programs	adjustments
tests of hypothesis	experimental design	

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The investigation considers aspects of the summarization, graphical display, and statistical analysis of data arising from chronic aquatic tests of toxicity with Daphnia magna and statistical issues involved in the planning of such tests. All procedures are illustrated with examples based on real data.

The report consists of fifteen sections. Section I describes the data and the toxicity tests from which they come. Sections II-VIII discuss and

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

illustrate various preliminary data displays and statistical analyses including preliminary graphical displays, tests for beaker to beaker heterogeneity within treatment groups, adjustments to account for such beaker to beaker heterogeneity, outlier detection procedure, comparisons of average response levels in teh water and solvent control groups, and the extent of association between length and reproduction responses. Sections IX-XV discuss hypothesis tests, multiple comparison procedures, and pairwise treatment group-control group confidence interval comparisons to detect treatment effects on mortality, reproduction, and length. These sections also discuss dose response curve fitting procedures to estimate those concentration levels that produce specified increases or decreases in response levels relative to the control group responses and growth curve analyses, analyses of time to death, and various statistical aspects of the design of toxicity tests.

AD _____

DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH *DAPHNIA MAGNA*

Final Report

by

Paul I. Feder
Applied Statistics Group, Battelle Columbus Laboratories

December, 1981

Supported by

U.S. Army Medical Research and Development Command
Fort Detrick, Frederick, Maryland 21701

Contract No. DAMD17-80-C-0165

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

Contract Officer's Technical Representative: Paul H. Gibbs
U.S. Army Medical Bioengineering Research and Development Laboratory
Fort Detrick, Frederick, Maryland 21701

EXECUTIVE SUMMARY

PURPOSE

During the past several decades problems of environmental contamination have become increasingly important, both from the scientific and the legal standpoints. In recent years a great deal of attention has been directed to the potential toxicity to acquatic organisms of chemicals discharged into water bodies.

The U.S. Army, through activities such as munitions manufacture, operates a number of plants that produce, consume, or discharge a variety of chemical substances. Some of these discharges enter bodies of water inhabited by various aquatic species. Thus the Army must provide the USEPA with safety data concerning the levels of such discharges and the possible extent of resulting suface water and ecological contamination. In order to develop such data the Army conducts both intramural and extramural programs of aquatic toxicity testing.

Considerable amounts of time, money, and manpower are expended by the Army in such aquatic toxicity testing programs. To make these programs more efficient and more effective, the need has been felt for a reexamination of some of the standard methods used. This has been especially true of statistical methods involved in the design of testing programs and the analysis of resulting data. Feder and Collins [1] have considered a number of the statistical aspects of the design and analysis of chronic aquatic toxicity tests with fathead minnows. The present study considers statistical aspects of the design and analysis of chronic tests with Daphnia magna.

Many of the statistical methods used in this study were adapted from those used by Feder and Collins [1] in their analysis of data from fathead minnow chronic toxicity tests. Other statistical methods discussed and/or developed in this study do not appear in [1]; some have not been previously applied to aquatic toxicity data. The methods discussed in [1] and in this report provide increased information, as compared with standard methodology, about the structure, relations, and anomolies in the data. Thus the statistical design and analysis considerations for Daphnia magna tests discussed in this report, in conjunction with those discussed by Feder and Collins for fathead minnow tests should improve the design, reporting, and statistical analysis of data from chronic aquatic toxicity tests. This will enhance the sensitivity of conclusions that can be derived from these tests, thereby increasing their efficiency.

APPROACH

The approach used in this report is much like that used by Feder and Collins [1]. A variety of topics pertaining to data display, statistical analyses, and experimental design are discussed in detail. For many of the topics, alternative statistical approaches and procedures are presented. All the statistical procedures are illustrated with examples based on real data from chronic tests with Daphnia magna. The data were kindly provided by several investigators at different laboratories and represent a number of variations in the design of chronic Daphnia tests.

The statistical procedures discussed in the body of the report represent a combination of methods that have been previously applied to aquatic toxicity data, methods that are in the statistical literature but which have not been commonly applied to aquatic toxicity data, and methods that have been especially developed or extended for this study.

RESULTS

A number of procedures for the statistical analysis of Daphnia magna chronic toxicity are discussed. The suggested procedures are illustrated with examples based on various chronic toxicity tests. The data used for illustration reflect a number of the possible variants in test design. Namely some of the tests are flow-through while others are static. Some contain multiple daphnids per beaker while others contain individual daphnids per beaker. Some contain solvent control groups while others do not. Numbers of daphnids per group vary from ten to eighty while numbers of beakers per group vary from three to ten.

Data analysis topics discussed include preliminary and residual graphical displays; preliminary tests of beaker to beaker heterogeneity within groups for mortality and length responses; adjustments to account for such heterogeneity; outlier detection tests; comparisons of average mortality, length, and reproduction levels between water control and solvent control groups along with conceptual implications of discrepancies that might be found; overall tests of heterogeneity in response levels across treatment groups; treatment group-control group pairwise multiple comparison and confidence interval procedures; the fitting of dose response curve models to mortality, reproduction, and length responses; point and confidence interval estimation, based on these fits, of concentrations associated with biologically significant increases or decreases in response levels relative to the controls; statistical precision to be expected from tests as a function of sample sizes, variability of responses, and extent of beaker to beaker heterogeneity within groups; a rationale for unequal allocation of test beakers within treatment groups, with greater numbers of beakers in the control group and lower treatment groups and lesser numbers of beakers in the higher treatment groups; analysis of time trends in reproduction; analysis of time to death.

CONCLUSIONS AND RECOMMENDATIONS

Several of the conclusions and recommendations from this study are similar to those arrived at by Feder and Collins [1]. We state some of them again here for completeness.

1. Standardized conventions and formats for reporting test design, laboratory conditions under which the test was carried out, and test results would facilitate communication among investigators, laboratories, and government regulatory agencies and would lead to greater reproducibility of test results across laboratories and across time.

2. Some of the "standard" methods currently used for analyzing data from aquatic toxicity tests can and should be modified. The data should first be graphed, outlying observations or groups of observations should be located and the reason for their aberrant behavior determined, and tests for heterogeneity among beakers within groups should be carried out. Subsequent comparisons of response levels across groups should take into account such heterogeneity or aberrant values.

3. Whenever solvent controls are included in the test their responses should be compared with those of the water controls. If no differences are evident, the two control groups may be combined for comparisons with the treatment groups. If there is evidence of differences among the groups then the solvent control group would usually be used for comparisons with the treatment groups. However the test results are then at best tentative since there is no way to determine whether the observed toxic effects were due to the toxicant, to the solvent, or to some interaction between them.

4. If hypothesis tests are to be used to compare the treatment group and control group responses they should be one sided tests which are sensitive to alternatives in a particular direction, rather than overall analysis of variance type "shotgun" tests.

5. Multiple comparison procedures and confidence intervals procedures should be used to determine specifically which treatment groups have responses which differ from the control group responses and whether the differences are of biological significance. Significance tests, by themselves, are not adequate to define an MATC*. Confidence bounds should be routinely constructed at the MATC to determine just how much worse than the control group the response at that concentration could conceivably be. In general, confidence intervals impart much more information than hypothesis tests and should be routinely used.

6. A way to impose monotonicity or smoothness structure on the responses, to smooth the data, and to convert a hypothesis testing problem into an estimation problem is to fit dose response curve models to the data and to define the "safe"concentration as that which results in no more than a specified increment in response

*Maximum acceptable toxicity conccentration.

from the control group. Dose response curves for mortality responses may be based on standard probit or logit models or nonstandard generalizations of these; dose response curves for length, reproduction, weight, or other quantitative responses may be based on multiple regression models such as polynomials or mechanistically motivated nonlinear forms.

7. Statistical power and precision depend on the number of daphnids per group, the number of beakers per group, and the extent of beaker to beaker and daphnid to daphnid response variability. In the presence of substantial beaker to beaker heterogeneity, the effective sample size per group may be closer to the number of beakers than to the number of daphnids. It is thus good design practice to divide the daphnids within each treatment or control group among as many beakers as can be accomodated within cost and logistical constraints.

8. Under certain circumstances it is sensible to allocate experimental resources so that the control group and lower concentration groups receive more beakers and daphnids than the higher concentration groups. This may result in greater inference sensitivity in the region of the MATC. Proportional diluters now have the capability to permit such asymmetrical allocations.

9. Statistical power or statistical precision goals should be stated as part of the protocol for each individual toxicity test and sample sizes should be determined accordingly.

10. Useful information can be obtained by studying concentration related trends in growth of reproduction with time and in time to death. Since these two responses are routinely determined under standard protocols at least three times per week, they should be reported at the intervals at which they're observed, to permit statistical analysis of the above responses. Costs premitting, consideration should be given to reporting reproduction and mortality on a more frequent basis than three times per week; perhaps daily.

PUBLICATIONS AND PRESENTATIONS

PAPERS

1. Feder, P. I., "Design and Analysis of Chronic Aquatic Tests of Toxicity With Daphnia Magna", October, 1981. Unpublished preprint.

PRESENTATIONS

1. "Design and Analysis of Aquatic Tests of Toxicity With Daphnia magna", Feder, P. I., Sixth Symposium on Aquatic Toxicology, American Society for Testing and Materials, Committee E-47, October, 1981.

PERSONNEL RECEIVING SUPPORT UNDER THE CONTRACT

Paul I. Feder

M. Claire Matthews

Lesley Arnold

David Bean

FOREWORD

I thank the following individuals and gratefully acknowledge their contributions to the project: Bill Adams, Gary Chapman, Clyde Goulden, and Gerald LeBlanc provided data sets to be used to illustrate the statistical methodology discussed in the report. These data sets represent important variations in the design of Daphnia toxicity tests and contain many interesting relationships. I appreciate the willingness of the investigators to share the basic results of their tests; William van der Schalie reviewed the data sets and the material describing the test conditions and enumerated them in a standardized and systematic form. This greatly helped me to understand the relationships of the different test designs relative to one another. David Bean, Gary Chapman, and Charles Stephan patiently helped me to understand the toxicological implications of a number of the statistical relationships observed in the data. Lesly Arnold and Claire Matthews organized and executed the data management aspects of the project and computed many of the statistical analyses and data displays contained in the report. Claire Matthews wrote the computer program, discussed in Section XII, that calculates point estimates and confidence intervals on "safe" concentrations. Paul Gibbs, the contract officer's technical representative, provided helpful cooperation, suggestions, and coordination throughout the study. Cynthia Wren did a fine job of typing the report.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF TABLES

INTRODUCTION

During the 1960's and 1970's, environmental contamination in general and water pollution specifically, became increasingly important as legal and scientific problems. Regulatory agencies needed scientific data to support the notion that a problem existed and also needed factual information for establishing tolerance limits for levels of chemical discharges into surface waters. From that need evolved numerous standard toxicity tests.

Aquatic toxicologists and biologists developed, refined, and standardized many of the biological, chemical, and operational factors pertaining to such tests. However the statistical aspects of test design and analysis of the resulting data have lagged behind.

Operational activities of the U.S. Army (e.g. munitions manufacture) involve the production, use, and/or discharge of a variety of commercial chemicals. Safety data must be provided to USEPA concerning surface water contamination due to discharges of chemical intermediates or the final product. The Army conducts both intramural and extramural programs of aquatic toxicity testing to develop such data.

Considerable amounts of time, money, and manpower are expended by the Army in such aquatic toxicity testing programs. To make these programs more efficient and more effective, the need has been felt for a reexamination of some of the standard methods used. This had been especially true of statistical methods involved in the design of testing programs and the analysis of resulting data.

This report represents the results of the second phase of a study of statistical methods in aquatic toxicology. The first phase pertained to chronic tests with fathead minnows; the results are presented in Feder and Collins [1]. The present study considers statistical aspects of the design and analysis of chronic tests with Daphnia magna.

Many of the statistical methods used in this study were adapted from those used by Feder and Collins [1] in their analysis of data from fathead minnow chronic toxicity tests. Other statistical methods discussed and/or developed in this study do not appear in [1]; some have not been previously applied to aquatic toxicity data and some have been especially developed or extended for this study. The suggested procedures are illustrated with examples based on various Daphnia chronic toxicity tests. The data used for illustration reflect a number of possible variants in test design.

Topics discussed include preliminary and residual graphical displays; preliminary tests of beaker to beaker heterogeneity within groups; adjustments to account for such heterogeneity; outlier detection tests; comparisons of average mortality, length, and reproduction levels between water control and solvent control groups along with conceptual implications of discrepanies that might be found; overall tests of heterogeneity in

response levels across treatment groups; treatment group-control group pairwise multiple comparison and confidence interval procedures; the fitting of dose response curve models to mortality, reproduction, and length responses; point and confidence interval estimation, based on these fits, of concentrations associated with biologically significant increases or decreases in response levels relative to the controls; statistical precision to be expected from tests as a function of sample sizes, variability of responses, and extent of beaker to beaker heterogeneity within groups; a rationale for unequal allocation of test beakers within treatment groups, with greater numbers of beakers in the control group and lower treatment groups and lesser numbers of beakers in the higher treatment groups; analysis of time trends in reproduction; analysis of time to death.

It is hoped that the results obtained in this study will contribute to better, more reliable toxicity tests and data analyses. This is turn should provide improved tools for the regulation of toxic chemicals in aquatic environments and should suggest fertile areas for further study and development.

I. CHRONIC DAPHNID TOXICITY TESTS - BASIC DATA

The data that we will be concerned with pertain to 21 day chronic Daphnia studies. There are a number of variations across laboratories in the way the tests were carried out and the types of responses that were recorded.

Some of the variations in technique do not affect statistical analysis of the data, although they are very important from a biological standpoint. Dr. William van der Schalie has compiled a collection of these biological parameters, which we indicate below in Table I.1.

Other test parameters strongly affect the responses that can be measured and the kinds of data analyses that can be carried out. Among these are:

- Number of replicate beakers (test chambers) per treatment level
- Number of daphnids per replicate beaker (test chamber)
  - individual daphnids
  - multiple daphnids
- Randomization procedures
  - test chamber to position
    - random, systematic, blocking
  - daphnids to test chambers
    - complete randomization
    - blocking - e.g. on broods
- Responses measured and frequency measured
  - live/dead - measured frequently or only once or twice during test
  - numbers of offspring produced - (total offspring or live offspring) — measured frequently or only once or twice during test
    - per replicate
    - per female
  - lengths
- Number of control groups
  - water control
  - solvent or carrier control

TABLE I.1 BIOLOGICAL, PHYSICAL, AND CHEMICAL SUMMARY PARAMETERS OF CHRONIC DAPHNIA TOXICITY TESTS (COMPILED BY DR. WILLIAM VAN DER SCHALIE)

Test Data Source:

Toxicant (Name or Code):

Daphnid Chronic Toxicity Test Data Sheet

Culture Information

Food Composition:

Frequency of Feeding:

Water Flow (static renewal, or flow-through):

Indicate frequency of renewal or flow rate in tank volumes/day:

Age of Adults used to obtain test Daphnids:

Temperature (average and range):

Photoperiod and Lighting (quality and intensity):

Was culture water used the same as for the test dilution water?

If not, how did the water quality differ:

Test Information (If same as for brood culture, please enter "same")

Food Composition:

Frequency of Feeding:

Water Flow (static renewal, or flow-through):

Indicate frequency of renewal of flow rate in tank volume/day:

Age of Daphnids at start of test:

Dilution Water Quality (indicate average and range during test, if possible):

Source (tap, well, river, reconstituted, etc):

pH: Conducitivity:

Alkalinity: Temperature:

Hardness: Other:

Dissolved oxygen:

TABLE I.1. (Continued)

Duration of Test:

Type of Test Container and Volume:

Nominal Concentrations:

Dilution Factor Between Concentrations:

Duration of Test:

Type of Test Container and Volume:

Number of Treatment Levels:

Number of Replicates per Treatment Level:

Number of Daphnids per Replicate:

Describe randomization procedures (daphnids to test containers and test containers to position):

Indicate which of the following biological endpoints were measured and the frequency measurement:

Survival:

Growth:

Days to first young:

Young production:

Total young per replicate:

Young per female:

Young per female per reproductive day:

Young per brood:

Dead or aborted young:

Other:

Control responses (indicate to 21 or 28 days)

Mortality ( ):

Young per female:

Were ephippia formed:

The data sets we have received represent a number of variations on the types of tests that are run. We have received five data sets.

1, 2. Gerald LeBlanc - EG&G Bionomics - Compounds "A" and "B"

Flow-through test - 21 day test
water control group, solvent control group, 5 treatment groups
4 replicate test chambers per group
20 daphnids per beaker to start
Responses measured
- 7, 14, 21 day survival
- 7, 14, 21 day cumulative offspring per surviving female
- 7, 21 day lengths (on individual daphnids)

Concentrations determined in two of four test chambers on days 0, 7, 14, 21 - Average concentration determinations used for analysis

3. Bill Adams - Selenium

Static renewal test - 21 day test
1 control group, 7 treatment groups
3 replicate test chambers per group
5 daphnids per beaker to start
Responses measured
- 2, 4, 7, 9, 11, 14, 16, 18, 21 day survival
- 2, 4, 7, 9, 11, 14, 16, 18, 21 day cumulative offspring per surviving female
- no length data

Averages of measured concentrations will be used for analysis

4. Gary Chapman - Beryllium

Static renewal test - 21 day test
water control group, solvent control group, 6 treatment groups
10 replicate test chambers per group
individual daphnids per test chamber
Responses measured
- 3, 5, 7, 10, 12, 14, 17, 19, 21 day survival
- 3, 5, 7, 10, 12, 14, 17, 19, 21 day # offspring (both live and dead) and # broods
- 21 day lengths

Averages of measured concentrations used for analysis

5. Clyde Goulden - Isophorone

Static renewal test - 21 day test
control group, 5 treatment groups
10 test chambers per group
- 7 beakers with individual daphnids - to determine individual production figures
- 3 beakers with 5 daphnids per beaker - to get survival information

Responses measured
- 2, 4, 6, 8, 11, 13, 15, 18, 21 day survival data

2, 4, 6, 8, 11, 13, 15, 18, 21 day # offspring (live)
from the daphnids in the beakers with just one daphnid per beaker
no length data
Nominal concentrations used for analysis

Note on Concentrations Used for Analysis Purposes

The various control groups and treatment groups correspond to nominal toxicant concentrations. As part of good experimental practice, periodic determinations are made of the chemical concentrations in the various beakers. The variation in these determinations is due in part to fluctuations over time in toxicant concentrations within the beakers and in part to analytical errors. Furthermore there may be variation in concentration levels among test chambers within groups due to random variations in the delivery system (e.g. partially blocked tubes) or varying amounts of settling out of toxicant in the various test chambers. However chemical determinations are not made in all chambers at equal intervals. We adopt the (somewhat arbitrary) conventions that:

1. If no chemical determinations are made during the experiment or if the results are not reported, we utilize the nominal concentrations in subsequent statistical analyses.

2. If chemical determinations are made within each group in representative test chambers at periodic intervals, then we associate the average of all these determinations with the concentration in each beaker within the group. That is, we utilize a single toxicant concentration over time and across beakers within groups.

   In flow through tests, theoretically all beakers within a group receive the same water and so should have the same concentrations. In static tests, this may not be so.

The problem of how to account for fluctuations in toxicant concentrations in the statistical analysis is an interesting one, however we will not pursue it here.

Different investigators and different laboratories report their data in different formats, in different styles, and at varying intervals. For example LeBlanc reports survival and production at days 7, 14, 21 while Adams, Chapman, and Goulden report them at more frequent intervals. LeBlanc and Chapman report lengths while Adams and Goulden do not. Some investigators report cumulative production versus time while other investigators report current production versus time. Some investigators report measured toxicant concentrations while others report only nominal concentrations.

Recommendation: A standardized data reporting format should be adopted, analogous to that discussed for the fathead minnow tests (Feder and Collins). A good start in that direction is the experimental categorization summary sheets prepared by Dr. William van der Schalie, which is shown in Table I.1.

In order to carry out statistical analyses and data displays, the data first had to be computerized. As the data were reported in a number of different formats, the data first needed to be recoded in a somewhat uniform manner, amenable for analysis. The approach we took is indicated in the figures below, which show computer listings of the various data sets.

Figures I.1 to I.4 are based on LeBlanc's tests on Compounds A and B. Figures I.1, I.3 contain 7, 14, 21 day survival and cumulative productivity, and day 0, 7, 14, 21 measured toxicant concentrations. Note that the concentrations given in the third field are nominal concentrations. Note also that since only two concentration determinations per group were made on each occasion, the "measured" concentrations on each day in each group are really two duplicates rather than four measured values. Figures I.2, I.4 contain seven day and 21 day length determinations on surviving adults at that point. (Note the 21 day length in beaker 7A but no corresponding 7 day length. This value should be associated with beaker C.)

Figure I.5 pertains to Adams' test on selenium. The concentrations indicated are nominal concentrations. Survival data is given corresponding to days 2, 4, 7, 9, 11, 14, 16, 18, 21. Cumulative fertility per surviving adult is indicated for these same days.

Figure I.6 pertains to Chapman's test on beryllium. Concentrations indicated are averages of measured concentrations for each treatment group. Since just one daphnid is contained in each test chamber, various responses can be measured that are not feasible with multiple daphnids per beaker, as in the LeBlanc or Adams tests. Thus time to death of each daphnid (50 represents a censored value--i.e. survives beyond day 21), number of broods, time to first brood, numbers of live young on days 3, 5, 7, 10, 12, 14, 17, 19, 21, and lengths are given. Note that in contrast to the LeBlanc and Adams productivity data, these productivity values represent current rather than cumulative values.

Figures I.7 and I.8 pertain to Goulden's test on isophorone. The concentrations indicated are nominal concentrations. Figure I.7 contains survival data for each of the test chambers on days 0, 2, 4, 6, 8, 11, 13, 15, 18, 21. Figure I.8 contains production data for the daphnids in beakers 1-7 within each group (individual daphnids per beaker). Time to death (50 represents a censored value), number of broods, time to first brood, and numbers of live young on days 2, 4, 6, 8, 11, 13, 15, 18, 21 are also given. As with the Chapman data, the production figures represent current rather than cumulative values.

INVESTIGATOR: EG AND G BIONOMICS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,1X,F6.4,4I5,3F5.1,4(A1,F6.4))

GRP	REPL	CONC (MG/L)	NO. TESTED		NO. LIVE (7,14,21)		CUM. FERT. (7,14,21)			MEASURED CONC (0,7,14,21)		
1	A	0.0000	20	20	17	17	2.0	48.0	84.0	0.00140	0.00069	0.00120	0.00068
1	B	0.0000	20	20	17	17	0.0	38.0	94.0	0.00140	0.00069	0.00120	0.00068
1	C	0.0000	20	20	18	17	1.0	82.0	129.0	0.00170	0.00063	0.00120	0.00072
1	D	0.0000	20	20	17	17	3.0	49.0	103.0	0.00170	0.00063	0.00120	0.00072
2	A	0.0000	20	20	20	18	1.0	96.0	164.0	0.00050	0.00065	0.00073	0.00090
2	B	0.0000	20	20	20	20	7.0	87.0	158.0	0.00050	0.00065	0.00073	0.00090
2	C	0.0000	20	20	20	20	1.0	70.0	134.0	0.00060	0.00059	0.00051	0.00058
2	D	0.0000	20	20	19	19	6.0	62.0	160.0	0.00060	0.00059	0.00051	0.00058
3	A	0.0031	20	20	20	20	3.0	62.0	149.0	0.00330	0.00330	0.00310	0.00210
3	B	0.0031	20	20	19	19	2.0	48.0	119.0	0.00330	0.00330	0.00310	0.00210
3	C	0.0031	20	20	15	15	3.0	30.0	120.0	0.00280	0.00310	0.00330	0.00250
3	D	0.0031	20	20	16	16	0.0	49.0	125.0	0.00280	0.00310	0.00330	0.00250
4	A	0.0062	20	20	19	17	1.0	67.0	153.0	0.00440	0.00430	0.00420	0.00130
4	B	0.0062	20	20	20	20	5.0	67.0	145.0	0.00440	0.00430	0.00420	0.00130
4	C	0.0062	20	20	19	17	0.0	52.0	113.0	0.00080	0.00420	0.00260	0.00110
4	D	0.0062	20	20	19	19	0.0	50.0	105.0	0.00080	0.00420	0.00250	0.00110
5	A	0.0120	20	20	20	18	7.0	77.0	189.0	0.00900	0.00680	0.00780	0.00470
5	B	0.0120	20	20	20	20	0.0	37.0	107.0	0.00900	0.00680	0.00780	0.00470
5	C	0.0120	20	20	18	16	0.0	35.0	112.0	0.00960	0.00880	0.00840	0.00420
5	D	0.0120	20	20	20	19	3.0	70.0	151.0	0.00960	0.00880	0.00840	0.00420
6	A	0.0250	20	20	19	18	0.0	30.0	103.0	0.01800	0.01800	0.01700	0.00770
6	B	0.0250	20	17	16	16	0.0	19.0	81.0	0.01800	0.01800	0.01700	0.00770
6	C	0.0250	20	7	6	5	2.0	35.0	91.0	0.0180	0.01500	0.01900	0.00800
6	D	0.0250	20	4	1	1	2.0	69.0	130.0	0.0180	0.01500	0.01900	0.00800
7	A	0.050	20	0	0	0	0.0	0.0	0.0	0.04100	0.03200	0.04100	0.01500
7	B	0.050	20	0	0	0	0.0	0.0	0.0	0.04100	0.03200	0.04100	0.01500
7	C	0.050	20	1	1	1	0.0	0.0	0.0	0.04000	0.00620	0.03900	0.01700
7	D	0.050	20	1	1	0	0.0	0.0	0.0	0.04000	0.00620	0.03900	0.01700

Figure I.1. LeBlanc - Test A. Survival, cumulative fertility, measured toxicant concentrations

```
INVESTIGATOR: E G AND G BIONOMICS - TEST A
GROWTH DATA FOR DAPHNIA AT DAY 7 AND DAY 21
GROUPS: CONTROL (1), SOLVENT CONTROL (2), TREATMENT (3-7)
4 BEAKERS (REPLICATIONS) PER GROUP
TEST, DAY, GROUP, BEAKER, FLEA MEASUREMENTS (UP TO 20)
FORMAT: (A1,I2,1X,I1,2X,20 F4.1)
A07 1 A  3.5 3.6 3.2 3.2 3.1 4.0 3.5 3.5 3.7 3.1 3.6 3.7 3.5 3.5 4.0 3.1 3.6 3.2 3.4 3.4
A07 1 B  3.0 3.5 3.3 2.9 3.4 3.0 3.5 3.6 3.6 3.5 3.7 3.1 3.2 3.4 3.0 3.3 3.2 3.4 2.6 2.9
A07 1 C  3.5 2.8 4.5 3.7 2.7 3.5 3.8 3.6 4.0 3.1 4.1 3.6 3.5 3.9 3.9 4.0 3.5 4.2 4.1 3.8
A07 1 D  3.5 4.0 3.1 3.2 3.2 3.7 3.6 3.1 4.0 3.5 3.5 3.3 3.1 3.6 3.4 3.1 3.6 3.2 3.5 3.4
A07 2 A  3.8 4.1 4.0 4.0 4.0 4.0 4.0 3.5 4.1 4.0 3.9 4.3 3.5 3.9 4.0 3.6 3.9 4.4 3.5 4.0
A07 2 B  3.5 3.6 4.0 3.6 4.0 3.4 4.1 3.9 3.8 3.5 3.4 3.6 4.0 3.9 3.8 3.6 3.1 3.4 3.5 3.8
A07 2 C  3.5 3.6 3.8 4.1 3.4 3.7 4.0 3.9 4.0 3.6 3.8 3.2 4.1 3.9 4.0 4.0 3.9 3.6 3.6 3.5
A07 2 D  3.8 3.6 4.0 3.4 4.0 4.0 4.1 3.6 3.9 3.2 3.9 4.1 3.9 4.0 3.6 3.9 4.4 3.5 3.5 3.4
A07 3 A  3.1 3.5 3.5 3.7 3.4 3.7 2.6 2.9 3.3 3.1 3.6 3.2 3.3 2.2 3.6 3.5 3.0 3.3 3.3 3.2
A07 3 B  4.0 2.5 3.9 3.4 2.2 3.7 4.0 3.9 3.0 3.8 3.5 3.2 2.2 2.5 3.5 2.2 2.9 3.0 2.5 4.0
A07 3 C  3.3 3.6 3.0 3.2 3.7 3.3 3.3 2.8 3.1 3.4 3.5 3.2 2.5 3.5 3.3 3.2 3.0 3.0 3.3 3.5
A07 3 D  3.4 3.1 3.4 3.5 3.4 3.7 3.5 3.7 2.6 3.5 3.4 3.0 3.1 3.2 3.5 3.8 3.5 2.9 3.5 3.5
A07 4 A  3.8 4.0 3.9 3.6 3.0 2.9 3.3 3.5 3.4 3.4 3.5 3.3 2.9 3.5 3.6 3.2 3.6 3.7 3.5 3.5
A07 4 B  3.5 3.5 3.3 3.2 3.6 3.8 3.6 3.6 3.2 3.4 3.5 3.5 3.5 3.5 3.2 3.2 3.0 3.4 3.7 3.8
A07 4 C  3.6 3.5 3.9 3.5 3.5 3.6 3.6 3.5 3.2 3.8 3.3 3.6 3.2 3.0 3.3 3.0 3.2 3.6 3.4 3.5
A07 4 D  3.4 3.3 3.7 3.5 3.5 3.5 3.5 3.5 2.6 3.2 3.4 3.7 3.6 3.7 3.5 3.7 2.9 3.5 3.6 3.3
A07 5 A  3.3 3.5 3.6 3.0 2.9 2.7 3.2 3.3 3.3 3.8 2.8 3.5 3.4 3.3 3.2 3.4 3.2 3.2 3.4 3.3
A07 5 B  2.9 3.3 3.2 3.0 3.3 3.4 3.3 3.9 3.8 2.8 3.5 3.5 3.5 3.4 3.2 3.2 3.4 3.7 3.4 3.5
A07 5 C  3.5 2.6 3.2 3.4 3.4 3.0 3.4 2.9 3.8 3.4 2.9 3.4 2.5 3.6 3.5 3.3 3.0 2.9 3.0 3.3
A07 5 D  3.5 3.2 2.7 3.4 3.0 3.8 3.1 3.1 3.1 3.4 3.7 3.6 3.3 3.3 2.7 3.5 3.3 3.6 3.0 3.5
A07 6 A  3.5 2.5 3.2 3.4 3.7 3.1 2.3 3.0 2.8 3.1 2.1 3.0 2.2 2.9 3.4 3.5 3.5 3.3 3.6 3.1
A07 6 B  2.9 2.9 2.8 3.0 3.0 3.2 2.3 3.2 2.9 3.2 3.2 3.4 3.1 3.1 3.4 3.2 3.2
A07 6 C  2.9 2.8 3.0 2.3 3.2 3.4 3.1
A07 6 D  3.4 3.1 2.3 3.1
A07 7 A
A07 7 B
A07 7 C  3.2
A07 7 D  3.5
A21 1 A  4.7 4.3 5.1 5.0 4.8 4.2 5.1 3.9 5.6 5.2 4.1 4.7 4.9 4.2 5.2 4.7 4.9
A21 1 B  5.2 5.3 5.1 [illegible] 5.0 5.6 5.5 4.8 5.1 5.0 5.5 5.0 4.6 5.0 5.0 5.0 5.0
A21 1 C  [illegible] 5.8 5.1 4.6 [illegible] 5.5 4.1 5.4 5.3 4.8 5.1 5.5 5.1 5.1 5.1 5.1 4.6
A21 1 D  4.8 4.9 [illegible] 4.6 4.8 5.5 4.8 4.9 4.5 5.0 5.0 5.1 4.4 4.5 4.9 4.9 4.9
A21 2 A  5.5 5.0 5.1 5.5 5.0 5.6 5.5 5.7 5.2 5.1 5.5 5.4 5.6 5.5 5.1 5.5 5.4 5.1
A21 2 B  5.0 5.5 5.2 5.5 5.2 5.1 5.0 5.4 4.5 4.9 4.6 5.5 5.6 5.3 5.6 5.2 5.3 5.2 5.5 5.5
A21 2 C  5.2 5.4 [illegible] [illegible] 5.2 [illegible] 5.0 5.0 5.8 4.5 5.8 5.5 4.9 5.1 5.1 4.9 5.2 4.8 4.8 5.5
A21 2 D  [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] 5.1 5.4 5.3 5.3 5.2 5.3 5.5 5.5 5.3 5.0 4.5 5.7 5.0
A21 3 A  [illegible] 5.0 5.2 5.2 5.0 5.2 5.2 5.0 5.2 5.0 5.6 5.5 5.2 5.2
A21 3 B  [illegible] 4.8 4.9 5.7 4.6 4.9 4.5 4.6 5.7 4.5 4.5 4.7 5.1
A21 3 C  [illegible] 4.6 [illegible] 4.8 5.0 4.7 4.2 5.0 4.2
A21 3 D  [illegible] 4.3 5.5 4.9 5.0 5.2 5.3 5.4 5.2 4.1
A21 4 A  [illegible] 4.8 4.6 5.2 5.6 4.8 5.3 5.3 5.1 5.4 5.5
A21 4 B  [illegible] 5.2 5.7 5.2 5.2 5.5 5.0 5.0 5.1 5.5 5.0 5.3 4.7 5.1
A21 4 C  [illegible] 4.5 4.8 5.0 5.3 4.9 4.6 5.3 5.1 4.8 5.1
A21 4 D  [illegible] 4.8 5.2 5.2 5.2 5.0 5.0 4.9 5.4 5.4 5.2 5.4 5.2
A21 5 A  [illegible] 5.6 [illegible] 5.1 5.0 5.1 5.3 5.3 4.5 5.3 5.5 5.6
A21 5 B  [illegible] 4.0 5.2 5.2 5.3 5.1 5.0 4.0 5.3 4.9 5.1 5.0 5.0 5.2
A21 5 C  [illegible] 4.1 4.6 4.9 4.7 5.0 4.8 4.5 5.3 5.1 4.9 5.0
A21 5 D  [illegible] 4.5 4.5 [illegible] 4.8 5.0 4.0 5.3 5.0 4.9 4.6 4.7 5.0 4.9
A21 6 A  [illegible] 4.6 5.5 4.8 4.0 5.9 5.0 5.5 4.8 5.3 4.7 5.5
A21 6 B  [illegible] 4.5 4.8 5.3 5.4 4.7 5.0 4.8 5.0 5.0 5.5
A21 6 C  [illegible]
A21 6 D  [illegible]
A21 7 A  [illegible]
A21 7 B
A21 7 C
A21 7 D
```

LeBlanc Test A. Seven day and 21 day lengths of surviving adults

INVESTIGATOR: EG AND G BIONOMICS--TEST B
21 DATA FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP - FORMAT: (I1,1X,A1,1X,F6.4,4I5,3F5.1,4(A1,F7.5)

GRP	REPL	CONC (MG/L)	NO. TESTED			NO. LIVE (7,14,21)		CUM. FERT. (7,14,21,)		MEASURED CONC (0,7,14,21)	
1	A	0.0000	20	20	20	19	0.0	27.0	79.0	-.99999	<0.00010 -.99999 <0.00020
1	B	0.0000	20	19	19	18	0.0	39.0	132.0	-.99999	<0.00010 -.99999 <0.00020
1	C	0.0000	20	20	20	20	0.0	41.0	127.0	<0.00010	-.99999 <0.00060 -.99999
1	D	0.0000	20	20	15	15	0.0	34.0	108.0	<0.00010	-.99999 <0.00060 -.99999
2	A	0.0000	20	19	18	17	0.0	33.0	106.0	-.99999	<0.00010 -.99999 <0.00020
2	B	0.0000	20	18	18	16	0.0	42.0	116.0	-.99999	<0.00010 -.99999 <0.00020
2	C	0.0000	20	19	18	16	0.0	25.0	83.0	<0.00010	-.99999 <0.00060 -.99999
2	D	0.0000	20	19	19	18	0.0	34.0	123.0	<0.00010	-.99999 <0.00060 -.99999
3	A	0.0120	20	20	20	20	0.0	42.0	169.0	-.99999	0.01300 -.99999 0.01500
3	B	0.0120	20	19	19	18	0.0	41.0	112.0	-.99999	0.01200 -.99999 0.01400
3	C	0.0120	20	20	19	20	0.0	26.0	77.0	0.01100	-.99999 0.02400 -.99999
3	D	0.0120	20	18	18	18	0.0	43.0	106.0	0.01200	-.99999 0.01400 -.99999
4	A	0.0250	20	20	20	17	0.0	57.0	155.0	-.99999	0.02800 -.99999 0.02200
4	B	0.0250	20	20	20	20	0.0	37.0	139.0	-.99999	0.03100 -.99999 0.02200
4	C	0.0250	20	20	20	20	0.0	41.0	117.0	0.04000	-.99999 0.02800 -.99999
4	D	0.0250	20	20	19	18	0.0	39.0	127.0	0.02500	-.99999 0.03200 -.99999
5	A	0.0500	20	20	19	18	0.0	40.0	119.0	-.99999	0.07300 -.99999 0.04600
5	B	0.0500	20	20	19	17	0.0	49.0	144.0	-.99999	0.07800 -.99999 0.04000
5	C	0.0500	20	20	19	15	0.0	45.0	135.0	0.04800	-.99999 0.05300 -.99999
5	D	0.0500	20	20	20	17	0.0	40.0	115.0	0.04400	-.99999 0.05100 -.99999
6	A	0.1000	20	20	20	18	0.0	44.0	122.0	-.99999	0.09500 -.99999 0.07400
6	B	0.1000	20	20	20	20	0.0	42.0	126.0	-.99999	0.07800 -.99999 0.09600
6	C	0.1000	20	20	20	16	0.0	45.0	107.0	0.09700	-.99999 0.03100 -.99999
6	D	0.1000	20	20	20	18	0.0	39.0	122.0	0.09900	-.99999 0.11000 -.99999
7	A	0.2000	20	5	5	3	0.0	30.0	110.0	-.99999	0.26000 -.99999 0.20000
7	B	0.2000	20	14	14	14	0.0	29.0	82.0	-.99999	0.22000 -.99999 0.21000
7	C	0.2000	20	12	12	12	0.0	31.0	114.0	0.19000	-.99999 0.35000 -.99999
7	D	0.2000	20	3	1	1	0.0	39.0	126.0	0.19000	-.99999 0.34000 -.99999

Figure 1.3. LeBlanc - Test B. Survival, cumulative fertility, measured toxicant concentrations

INVESTIGATOR: EG AND G BIONOMICS - TEST B
GROWTH DATA FOR DAPHNIA AT DAY 7 AND DAY 21
GROUPS: CONTROL (1), SOLVENT CONTROL (2), TREATMENT (3-7)
4 BEAKERS (REPLICATIONS) PER GROUP
TEST, DAY, GROUP, BEAKER, FLEA MEASUREMENTS (UP TO 20)
FORMAT: (A1,I2,1X,I1,2X,20F4.1)

```
B07 1 A  3.4 3.3 4.2 3.3 2.9 2.6 3.5 4.2 3.5 2.9 3.9 3.4 3.4 3.0 3.3 3.5 3.7 3.1 3.0 2.4
B07 1 B  3.2 3.8 3.5 4.2 3.1 2.8 3.7 3.5 3.7 3.6 3.1 3.5 3.4 3.4 3.9 3.5 3.7 3.6 3.4
B07 1 C  4.3 4.0 3.3 3.9 4.1 3.8 3.0 3.7 3.2 4.2 4.0 4.2 4.3 3.6 3.5 4.2 3.7 3.6 3.8 3.1
B07 1 D  3.2 3.4 3.5 3.6 3.1 3.9 3.0 4.5 3.6 4.3 3.8 3.5 3.2 3.8 3.6 3.5 3.6 4.0 2.9
B07 2 A  3.7 3.3 4.0 2.7 4.2 2.9 3.7 3.5 3.5 3.7 3.9 3.2 3.7 3.1 3.5 3.8 3.7 2.3 3.6
B07 2 B  2.8 3.9 3.8 3.5 2.8 3.4 3.8 3.5 3.8 3.5 3.4 3.4 3.4 2.9 3.8 4.0 3.8 3.3
B07 2 C  3.8 4.0 4.0 3.8 3.0 3.2 3.5 3.8 2.8 3.7 3.5 3.1 3.1 3.9 3.0 3.3 3.5 3.3 3.0
B07 2 D  2.0 3.6 4.2 3.6 3.3 3.6 3.9 3.8 3.5 3.5 3.6 3.9 4.0 3.5 3.6 3.5 3.7 3.6 3.8
B07 3 A  3.8 3.6 3.3 3.6 3.1 3.0 3.7 3.7 3.1 3.6 3.6 3.2 3.3 3.1 3.3 2.8 3.6 3.0 3.3 3.6
B07 3 B  3.6 4.0 3.6 3.5 4.3 3.4 3.8 4.4 4.0 3.9 4.3 3.2 3.2 3.8 3.8 3.5 3.5 3.5 3.3
B07 3 C  3.4 3.5 3.0 3.3 3.5 3.5 3.8 3.3 4.1 3.4 2.8 3.8 4.0 2.9 3.7 3.4 3.4 3.8 3.1 3.2
B07 3 D  4.5 3.8 3.3 3.1 4.0 3.2 3.8 2.7 3.0 3.8 4.0 4.2 3.4 3.6 3.4 3.5 3.5 3.1
B07 4 A  3.4 3.5 3.2 3.2 3.7 3.5 3.6 3.3 3.3 3.0 4.1 3.6 3.4 3.2 2.0 4.0 4.1 3.5 3.8 3.2
B07 4 B  3.6 3.5 3.3 3.0 2.8 2.5 3.9 4.0 3.8 3.8 2.8 3.5 2.8 3.6 3.6 3.4 3.4 4.0 2.8 3.2
B07 4 C  3.9 4.0 4.2 4.2 3.8 3.2 3.9 3.4 3.6 3.6 4.0 2.8 3.8 3.3 2.3 3.5 3.4 3.5 4.0 3.6
B07 4 D  4.0 3.5 3.5 3.6 4.1 3.2 4.1 3.6 3.8 3.6 3.8 3.6 4.0 3.6 3.5 3.5 3.3 3.5 3.8 3.8
B07 5 A  4.1 3.6 3.5 2.9 3.2 4.0 4.0 3.9 3.7 4.1 3.2 3.2 3.6 2.8 3.8 3.9 2.9 3.8 3.3 3.6
B07 5 B  4.0 3.4 4.2 3.2 3.9 2.7 3.8 3.5 3.7 4.0 2.6 3.0 4.1 3.7 3.8 3.5 3.7 3.8 3.2 3.6
B07 5 C  3.9 3.3 3.6 3.3 3.8 3.7 3.3 3.6 3.2 3.8 3.0 3.5 3.5 3.5 3.9 3.8 3.5 3.6 4.0 3.2
B07 5 D  2.8 3.6 3.3 3.7 3.0 3.2 3.6 3.8 3.2 3.0 3.5 2.9 3.5 4.4 4.0 3.2 3.8 3.2 4.0 3.8
B07 6 A  3.2 3.2 3.1 3.5 4.0 3.4 3.1 4.0 4.0 3.4 2.0 4.0 3.3 3.9 3.7 3.8 3.5 3.9 3.8 3.7
B07 6 B  3.9 3.6 3.3 3.6 4.0 3.7 2.8 3.2 3.7 3.8 3.3 3.1 3.6 3.8 3.0 3.2 4.0 4.0 3.0 3.8
B07 6 C  3.7 3.3 4.2 4.1 3.2 4.3 3.6 3.9 4.1 3.8 4.0 4.4 3.4 4.0 4.0 4.3
B07 6 D  3.8 4.0 4.0 3.7 3.2 3.5 3.7 3.8 3.9 4.3 3.8 3.7 3.7 3.2 3.1 3.6 4.0 3.5 4.0 4.0
B07 7 A  2.3 3.5 3.0 3.2 2.4
B07 7 B  3.2 3.5 4.5 3.3 3.5 2.8 3.0 3.0 2.8 2.8 3.2 3.0 3.0 3.3
B07 7 C  3.8 3.7 3.2 2.6 3.1 3.1 3.6 3.1 3.4 4.1 3.2 3.4
B07 7 D  3.0 2.5 1.6
B21 1 A  5.4 4.8 4.3 5.2 5.1 4.2 5.0 4.9 4.8 5.7 5.2 5.1 5.2 5.2 5.0 4.8 5.0 5.2 5.4
B21 1 B  4.9 4.9 4.9 5.0 5.2 5.2 5.0 4.5 4.9 4.9 5.3 4.5 4.7 4.9 5.5 4.6 4.8 4.7
B21 1 C  5.4 4.6 5.1 4.5 5.0 5.0 5.5 5.1 4.9 5.2 4.9 4.8 5.0 4.9 4.6 4.9 4.0 4.6 5.4 4.9
B21 1 D  4.7 4.8 5.1 4.7 5.0 5.0 4.6 5.2 5.0 4.8 5.0 5.2 5.2 5.4 4.7
B21 2 A  4.5 4.5 5.0 4.5 4.5 5.2 5.0 5.0 4.9 5.0 4.9 4.6 5.0 4.5 5.0 5.0 4.5
B21 2 B  5.2 5.4 4.5 4.6 4.6 4.5 4.5 4.9 4.7 5.2 5.2 4.2 4.8 4.8 4.6 4.3
B21 2 C  5.0 4.7 4.7 4.0 4.0 4.9 5.1 4.9 4.3 4.2 4.7 4.5 5.0 4.9 4.9 4.6
B21 2 D  5.2 5.5 5.0 5.2 5.1 5.1 5.2 4.4 5.5 5.0 5.3 5.0 5.6 5.2 4.1 4.4 5.4 5.0
B21 3 A  4.1 4.6 4.8 4.5 5.5 4.7 4.7 5.5 4.0 4.7 5.4 5.0 5.4 4.6 5.0 5.0 4.7 5.5 4.0 4.1
B21 3 B  4.6 4.5 4.7 4.8 5.1 5.2 5.0 5.2 5.8 4.5 5.5 5.5 4.8 5.0 5.6 4.6 4.5 5.0
B21 3 C  4.9 4.7 4.9 4.3 4.0 5.7 4.8 4.0 4.5 5.0 4.1 4.7 5.2 4.8 4.6 4.8 4.7 4.5 4.3 5.0
B21 3 D  4.9 5.1 5.5 5.1 5.1 4.7 5.0 4.4 4.8 4.8 4.8 4.1 5.8 5.7 5.8 4.9 4.5 5.5
B21 4 A  4.3 5.0 5.0 5.2 4.3 5.5 5.5 4.5 5.0 5.2 5.0 5.0 4.9 4.8 5.0 5.2 5.0
B21 4 B  4.8 5.7 4.8 4.8 4.9 5.3 4.6 5.5 4.9 4.8 4.8 4.8 5.5 5.5 5.0 5.3 5.0 4.6 4.8 5.0
B21 4 C  4.8 5.2 4.7 4.9 4.4 5.0 5.7 4.4 5.5 5.2 4.1 5.2 5.0 5.0 4.5 5.0 5.3 4.7 5.0 5.0
B21 4 D  4.2 4.3 5.0 5.5 4.8 4.8 4.5 5.0 5.5 5.5 5.3 4.7 4.8 4.9 4.3 4.7 4.4 4.6
B21 5 A  4.4 5.2 4.2 5.0 5.0 4.8 4.8 5.3 4.8 4.8 5.3 4.5 4.2 4.6 5.1 5.2 4.6 5.3
B21 5 B  4.2 5.0 5.0 5.0 4.5 4.7 4.4 5.0 4.4 5.0 4.9 4.9 4.9 4.9 4.9 5.0 5.0
B21 5 C  4.7 4.5 4.8 4.8 5.3 4.7 5.0 4.5 4.6 4.4 5.2 4.8 5.0 4.8 4.5
B21 5 D  5.5 5.1 5.1 5.1 4.1 5.1 4.8 5.0 4.4 5.2 4.6 4.6 5.1 5.1 5.0 4.5 5.3
B21 6 A  4.5 4.5 5.4 5.4 5.4 5.0 4.7 5.6 4.6 4.7 4.5 4.5 4.1 5.4 5.1 5.1 5.8 4.5
B21 6 B  4.7 5.0 5.0 4.3 4.2 4.8 4.5 4.0 5.3 4.7 4.6 4.6 4.7 4.5 4.8 4.3 4.2 4.6 5.1 4.7
B21 6 C  5.3 5.5 4.6 4.5 4.6 4.4 4.9 4.7 4.1 5.1 4.6 4.6 5.2 5.2 4.5 5.1
B21 6 D  5.3 5.3 5.2 5.2 4.4 5.2 5.0 5.4 5.0 4.8 5.3 4.2 5.3 5.3 4.2 5.2 4.4 5.0
B21 7 A  4.9 4.1 4.7
B21 7 B  4.3 4.9 4.6 5.3 4.5 4.6 4.9 4.6 4.8 5.0 4.6 5.1 4.7 4.1
B21 7 C  5.4 4.0 4.7 5.0 4.7 5.0 4.6 5.1 4.3 5.0 5.4 4.5
B21 7 D  4.4
```

Figure I.4. LeBlanc - Test B.

Seven day and 21 day lengths of surviving adults

INVESTIGATOR: WJ ADAMS TEST-SELENIUM-SURVIVORS,CUM.FERT/FLEA
STATIC RENEWAL TESTS WITH DAPHNIA MAGMA
1 CONTROL(1), 7 TRX GROUPS(2-8)/FORMAT:(I1,1X,A1,F5.0,10I2,10F6.1)
3 BEAKERS(REPL) PER GROUP---VARIABLES ARE
GRP,REPL,CONC,NO.TESTED,NO.ALIVE(2,4,7,9,11,14,16,18,21),
CUM.FERT/FEMALE(2,4,7,9,11,14,16,18,21)

1	A	0.01	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	2.0	19.6	35.0	35.8	54.0
1	B	0.01	5	5	5	5	5	5	4	4	4	0.0	0.0	0.0	0.0	0.0	13.3	23.0	27.5	40.0
1	C	0.01	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	4.0	20.6	31.4	36.6	49.2
2	A	.065	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	1.0	14.0	26.2	29.4	42.4
2	B	.065	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	3.4	17.4	26.6	29.4	43.8
2	C	.065	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	1.0	18.2	26.2	34.2	49.0
3	A	.115	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	5.4	14.8	29.6	33.8	50.0
3	B	.115	5	5	5	-9	5	5	5	5	5	0.0	0.0	0.0	0.0	0.0	7.2	13.2	14.0	26.6
3	C	.115	5	5	5	5	5	5	5	5	4	0.0	0.0	0.0	1.4	3.0	17.8	24.4	32.2	47.7
4	A	.235	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	0.0	5.8	12.6	19.2	28.4
4	B	.235	5	5	5	5	5	5	5	5	5	0.0	0.0	0.0	0.0	0.0	2.4	6.8	15.4	36.8
4	C	.235	5	5	-9	5	5	5	5	5	5	0.0	0.0	0.0	0.0	8.0	17.8	31.4	35.6	53.6
5	A	.555	5	5	2	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	B	.555	5	5	2	1	1	1	1	1	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	C	.555	5	5	1	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	A	0.91	5	5	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	B	0.91	5	4	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	C	0.91	5	5	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	A	2.00	5	1	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	B	2.00	5	1	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	C	2.00	5	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	A	4.00	5	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	B	4.00	5	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	C	4.00	5	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 1.5. Adams - Selenium - 21 day survival and production data

INVESTIGATOR: G. CHAPMAN- INDIVIDUAL DAPHNIA PER BEARER (BEP), 10 BEPS PER GROUP
STATIC RENEWAL TEST WITH DAPHNIA MAGNA: GROUPS: CONTROL(1), CARRIER CONTROL(2),
TREATMENT GROUPS(3-8), VARIABLES:
GRP, REP, MEASURED CONC., TIME TO DEATH, NO. BROODS, TIME TO FIRST BROOD,
NO. LIVE YOUNG (DAYS 1,5,7,10,12,14,17,19,21), LENGTH
FORMAT:(I1,I2,F6.3,3I4,2X,9I3,F5.1)

1 1	0.0	50	3	14	0	0	0	0	0	5	23	42	0	4.0
1 2	0.0	50	3	12	0	0	0	0	6	0	28	41	0	4.2
1 3	0.0	50	5	7	0	0	0	6	34	0	23	45	64	4.4
1 4	0.0	50	5	10	0	0	0	9	16	5	30	47	0	4.2
1 5	0.0	50	3	10	0	0	0	14	0	0	18	33	0	4.0
1 6	0.0	50	6	10	0	0	0	46	0	36	48	75	47	4.8
1 7	0.0	50	5	12	0	0	0	0	1	24	39	55	59	4.2
1 8	0.0	50	5	10	0	0	0	58	0	43	19	46	0	4.1
1 9	0.0	50	5	10	0	0	0	15	33	39	44	56	0	4.5
110	0.0	50	5	10	0	0	0	47	0	34	36	49	59	4.4
2 1	0.0	50	5	10	0	0	0	3	14	10	29	43	0	4.0
2 2	0.0	14	2	10	0	0	0	7	16	0	0	0	0	
2 3	0.0	3	0	-1	0	0	0	0	0	0	0	0	0	
2 4	0.0	50	6	10	0	0	0	1	17	6	21	44	60	4.2
2 5	0.0	50	5	10	0	0	0	45	0	46	44	56	0	4.3
2 6	0.0	50	5	10	0	0	0	18	10	38	30	51	0	4.1
2 7	0.0	50	6	10	0	0	0	45	0	21	29	58	47	4.3
2 8	0.0	50	5	10	0	0	0	52	0	15	28	62	38	4.6
2 9	0.0	50	5	10	0	0	0	5	37	52	53	52	40	4.4
210	0.0	50	6	10	0	0	0	52	0	25	31	57	35	4.3
3 1	2.60	3	0	-1	0	0	0	0	0	0	0	0	0	
3 2	2.60	50	5	10	0	0	0	4	24	0	21	0	34	4.1
3 3	2.60	10	0	-1	0	0	0	0	0	0	0	0	0	
3 4	2.60	50	5	10	0	0	0	45	0	7	9	36	0	4.4
3 5	2.60	50	4	10	0	0	0	8	12	0	10	0	21	3.9
3 6	2.60	50	3	12	0	0	0	0	3	0	7	20	0	3.8
3 7	2.60	50	5	10	0	0	0	37	0	11	17	42	44	4.3
3 8	2.6	50	4	10	0	0	0	2	13	0	9	38	0	4.1
3 9	2.6	50	5	10	0	0	0	9	14	8	25	50	0	4.2
310	2.6	50	5	10	0	0	0	12	7	13	37	57	0	4.2
4 1	6.21	12	2	7	0	0	0	16	0	0	0	0	0	
4 2	6.21	3	0	-1	0	0	0	0	0	0	0	0	0	
4 3	6.21	50	3	10	0	0	0	2	0	0	14	26	0	4.0
4 4	6.21	3	0	-1	0	0	0	0	0	0	0	0	0	
4 5	6.21	3	0	-1	0	0	0	0	0	0	0	0	0	
4 6	6.21	50	3	17	0	0	0	0	0	0	4	16	29	3.7
4 7	6.21	50	5	10	0	0	0	23	2	9	19	46	0	4.2
4 8	6.21	50	5	10	0	0	0	22	0	13	18	40	0	4.3
4 9	6.21	50	5	10	0	0	0	2	6	0	8	24	52	4.0
410	6.21	50	5	12	0	0	0	0	11	4	6	19	32	3.8
5 1	13.73	3	0	-1	0	0	0	0	0	0	0	0	0	
5 2	13.73	3	0	-1	0	0	0	0	0	0	0	0	0	
5 3	13.73	12	1	10	0	0	0	15	0	0	0	0	0	
5 4	13.73	50	4	10	0	0	0	5	29	0	13	29	0	3.8
5 5	13.73	10	1	10	0	0	0	2	0	0	0	0	0	
5 6	13.73	50	4	14	0	0	0	0	0	11	25	56	43	4.0
5 7	13.73	50	3	14	0	0	0	0	0	6	5	28	0	3.8
5 8	13.73	50	5	10	0	0	0	8	10	31	42	44	0	4.1
5 9	13.73	50	5	10	0	0	0	11	16	24	38	0	57	4.2
510	13.73	50	3	10	0	0	0	5	0	0	20	0	18	3.0
6 1	26.88	50	5	10	0	0	0	6	1	1	8	0	16	3.3
6 2	26.88	50	4	10	0	0	0	2	0	7	17	27	0	3.5
6 3	26.88	50	5	10	0	0	0	2	3	13	17	28	0	3.8
6 4	26.88	50	4	12	0	0	0	0	2	22	39	39	0	3.8
6 5	26.88	10	0	-1	0	0	0	0	0	0	0	0	0	
6 6	26.88	50	4	12	0	0	0	0	15	30	52	66	0	4.3
6 7	26.88	50	5	12	0	0	0	6	7	27	45	62	24	4.1
6 8	26.88	50	5	10	0	0	0	0	10	18	0	107	0	58	4.2
6 9	26.88	50	5	10	0	0	0	6	0	26	39	56	43	4.3
610	26.88	50	3	10	0	0	0	10	0	0	23	0	28	3.8
7 1	54.57	50	5	10	0	0	0	6	0	12	26	47	44	3.7
7 2	54.57	50	5	10	0	0	0	1	15	16	7	0	12	3.3
7 3	54.57	50	4	10	0	0	0	3	8	0	5	10	0	2.8
7 4	54.57	50	4	12	0	0	0	0	13	24	19	50	0	3.9
7 5	54.57	50	6	7	0	0	5	7	18	0	86	0	35	3.9
7 6	54.57	50	4	12	0	0	0	0	5	26	36	45	0	4.1
7 7	54.57	50	5	10	0	0	0	7	16	18	12	23	0	3.6
7 8	54.57	50	5	10	0	0	0	9	0	0	6	16	7	2.9
7 9	54.57	50	6	7	0	0	5	11	21	32	46	0	34	4.2
710	54.57	50	5	10	0	0	0	10	15	27	10	0	10	3.5
8 1	111.08	12	2	10	0	0	0	1	10	0	0	0	0	
8 2	111.08	50	5	10	0	0	0	5	0	11	24	23	22	3.3
8 3	111.08	50	5	10	0	0	0	1	10	14	16	25	0	3.4
8 4	111.08	50	4	12	0	0	0	0	14	20	28	41	0	3.9
8 5	111.08	50	5	10	0	0	0	3	20	20	26	45	0	3.9
8 6	111.08	14	2	12	0	0	0	0	13	9	0	0	0	
8 7	111.08	50	4	12	0	0	0	0	5	0	28	22	0	3.3
8 8	111.08	50	6	7	0	0	6	8	17	25	40	0	0	4.1
8 9	111.08	7	0	-1	0	0	0	0	0	0	0	0	0	
810	111.08	21	3	14	0	0	0	0	0	25	30	57	0	

Figure 1.6. Chapman - Beryllium.

Survival, production, and 21 day length data

1 1	0	1	1	1	1	1	1	1	1	1	1
1 2	0	1	1	1	1	1	1	1	1	1	1
1 3	0	1	1	1	1	1	1	1	1	1	1
1 4	0	1	1	1	1	1	1	1	1	1	1
1 5	0	1	1	1	1	1	1	1	1	1	1
1 6	0	1	1	1	1	1	1	1	1	1	1
1 7	0	1	1	1	1	1	1	1	1	1	1
1 8	0	5	5	5	5	5	5	5	4	4	4
1 9	0	5	5	5	5	5	5	5	5	5	4
110	0	5	5	5	5	5	5	5	5	5	5
2 1	10	1	1	1	1	1	1	1	1	1	0
2 2	10	1	1	1	1	1	1	1	1	1	1
2 3	10	1	1	1	1	1	1	1	1	1	1
2 4	10	1	1	1	1	1	1	1	1	1	1
2 5	10	1	1	1	1	1	1	1	1	1	1
2 6	10	1	1	1	1	1	1	1	1	1	1
2 7	10	1	1	1	1	1	1	1	1	1	1
2 8	10	5	5	5	5	5	5	5	4	4	4
2 9	10	5	5	5	5	5	5	5	5	5	5
210	10	5	5	5	5	5	5	5	5	5	5
3 1	50	1	1	1	1	1	1	1	1	1	1
3 2	50	1	1	1	1	1	1	1	1	1	1
3 3	50	1	1	1	1	1	1	1	1	1	1
3 4	50	1	1	1	1	1	1	1	1	1	1
3 5	50	1	1	1	1	1	1	1	1	1	1
3 6	50	1	1	1	1	1	1	1	1	1	1
3 7	50	1	1	1	1	1	1	1	1	1	1
3 8	50	5	5	5	5	5	5	5	4	4	4
3 9	50	5	5	5	5	5	5	5	5	5	5
310	50	5	5	5	5	5	5	5	5	5	5
4 1	100	1	1	1	1	1	1	1	1	1	1
4 2	100	1	1	1	1	1	1	1	1	1	1
4 3	100	1	1	1	1	1	1	1	1	1	1
4 4	100	1	1	1	1	1	1	1	1	1	1
4 5	100	1	1	1	1	1	1	1	1	1	1
4 6	100	1	1	1	1	1	1	1	1	1	1
4 7	100	1	1	1	1	1	1	1	1	1	1
4 8	100	5	5	4	4	4	4	4	4	4	3
4 9	100	5	5	5	5	5	5	5	4	4	4
410	100	5	5	5	5	5	5	5	5	5	5
5 1	150	1	1	1	1	1	1	1	1	1	1
5 2	150	1	1	1	1	1	1	1	1	1	1
5 3	150	1	1	1	1	1	1	1	1	1	1
5 4	150	1	1	1	1	1	1	1	1	1	1
5 5	150	1	1	1	1	1	1	1	1	1	1
5 6	150	1	1	1	1	1	1	1	1	1	1
5 7	150	1	1	1	1	1	1	1	1	1	1
5 8	150	5	5	4	4	4	1	1	0	0	0
5 9	150	5	5	5	5	5	5	5	4	3	3
510	150	5	5	5	5	5	5	5	5	4	4
6 1	200	1	1	1	1	1	0	0	0	0	0
6 2	200	1	1	1	1	1	0	0	0	0	0
6 3	200	1	1	1	1	1	0	0	0	0	0
6 4	200	1	1	1	1	1	0	0	0	0	0
6 5	200	1	1	1	1	1	1	1	1	1	1
6 6	200	1	1	1	1	0	0	0	0	0	0
6 7	200	1	1	1	1	1	0	0	0	0	0
6 8	200	5	5	5	5	4	3	1	1	1	1
6 9	200	5	5	5	5	4	2	1	1	1	1
610	200	5	5	5	1	0	0	0	0	0	0

Figure I.7. Goulden - Isophorone. Survival data

GOULDEN DATA--EXPLORATORY ANALYSES ON REPRODUCTION

GRP REP CONC TTD NOBROOD T1BROOD NLY1 TO NLY9

(F1.0,F2.0,F7.0,3F4.0,2X,9F3.0)

1	1	0	50	4	11	0	0	0	0	25	21	2	1	0
1	2	0	50	4	11	0	0	0	0	10	34	17	0	15
1	3	0	50	4	11	0	0	0	0	9	33	19	0	14
1	4	0	50	4	11	0	0	0	0	11	38	23	0	8
1	5	0	50	5	11	0	0	0	0	17	39	35	24	20
1	6	0	50	4	11	0	0	0	0	12	34	6	16	15
1	7	0	50	3	11	0	0	0	0	1	34	0	0	14
2	1	10	18	3	11	0	0	0	0	7	34	18	0	0
2	2	10	50	4	11	0	0	0	0	11	42	21	0	3
2	3	10	50	4	11	0	0	0	0	10	36	21	0	8
2	4	10	50	4	11	0	0	0	0	8	32	26	0	13
2	5	10	50	5	11	0	0	0	0	9	31	30	19	24
2	6	10	50	4	8	0	0	0	4	41	0	25	31	0
2	7	10	50	4	11	0	0	0	0	10	27	24	0	13
3	1	50	50	4	11	0	0	0	0	15	38	0	26	38
3	2	50	50	4	11	0	0	0	0	12	33	25	14	0
3	3	50	50	4	11	0	0	0	0	7	35	30	0	13
3	4	50	50	4	11	0	0	0	0	7	23	18	19	0
3	5	50	50	5	11	0	0	0	0	6	33	38	3	16
3	6	50	50	4	11	0	0	0	0	34	0	22	25	18
3	7	50	50	4	13	0	0	0	10	25	0	0	29	24
4	1	100	50	4	11	0	0	0	0	7	24	21	15	0
4	2	100	50	5	11	0	0	0	0	10	28	23	13	2
4	3	100	50	4	11	0	0	0	0	10	23	9	0	16
4	4	100	50	4	11	0	0	0	0	10	20	13	0	19
4	5	100	50	5	11	0	0	0	0	17	21	8	20	11
4	6	100	50	5	11	0	0	0	0	9	18	19	19	10
4	7	100	50	4	11	0	0	0	0	5	29	21	13	0
5	1	150	50	2	15	0	0	0	0	0	0	11	0	9
5	2	150	50	1	15	0	0	0	0	0	0	7	0	0
5	3	150	50	4	11	0	0	0	0	2	10	0	17	11
5	4	150	50	2	13	0	0	0	0	0	3	0	5	0
5	5	150	50	2	13	0	0	0	0	0	2	0	0	13
5	6	150	50	0	-1	0	0	0	0	0	0	0	0	0
5	7	150	50	1	13	0	0	0	0	0	3	0	0	0
6	1	200	11	0	-1	0	0	0	0	0	0	0	0	0
6	2	200	11	0	-1	0	0	0	0	0	0	0	0	0
6	3	200	11	0	-1	0	0	0	0	0	0	0	0	0
6	4	200	11	0	-1	0	0	0	0	0	0	0	0	0
6	5	200	50	0	-1	0	0	0	0	0	0	0	0	0
6	6	200	8	0	-1	0	0	0	0	0	0	0	0	0
6	7	200	11	0	-1	0	0	0	0	0	0	0	0	0

Figure I.8. Goulden - Isophorone. Time to death, production data

II. PRELIMINARY GRAPHICAL DATA DISPLAYS

Graphing the data is generally considered to be a good first step in statistical analysis. Graphs provide insights into the structure of the data and reveal the presence of possibly unanticipated relations or anomolies. This section contains preliminary displays of survival, production, and length responses versus group and versus concentration.

Figures II.1 to II.10 display trends in 21 day mortality (or survival) with group and with concentration for the LeBlanc, Adams, Chapman, and Goulden tests.

Figures II.1 to II.4a pertain to LeBlanc's tests with Compounds A and B respectively. These tests consist of a water control group, a solvent control group, and five treatment groups, with four replicate test chambers per group. Figures II.1, II.2, and II.2a show both a higher mortality level and a higher average measured concentration in the water control group than in the solvent control group for test A (15 percent versus 3.75 percent mortality and 0.0010 mg/ℓ versus 0.0006 mg/ℓ concentration respectively). There is no trend in mortality among the first three treatment groups (up to nominal concentration 0.0120 mg/ℓ and then a rather rapid rise as the nominal concentration level is doubled and then redoubled from that point. No outlying observations are evident, however there is some question of the homogeneity of responses across replicates in group 6. Figures II.3, II.4, and II.4aa show virtually no trend in mortality with increasing concentration in test B, up to group 6. There is then a substantial increase in mortality between group 6 and group 7. No outlying results are evident, but there is question about homogeneity of responses across replicate test chambers in group 7.

Figures II.5 and II.6 pertain to Adams' test with selenium. This test consists of a control group and seven treatment groups with three replicate test chambers per group. There appears to be no trend in mortality (and very little mortality) in groups 1 to 4. There is then a sudden jump in mortality between groups 4 and 5 and mortality remains at 100 percent thereafter. No outlying responses are evident. We conclude from these graphs that the dose response relation is very steep.

Figure II.7 pertains to Chapman's test with beryllium. The test consists of water and solvent control groups and six treatment groups. There are individual daphnids in each beaker. The mortality rates plotted represent average mortality over the ten beakers within each group. There appears to be no systematic trend in mortality with increasing concentration nor does there appear to be any outlying responses.

Figures II.8, II.9 and II.10 pertain to Goulden's test with isophorone. The test consists of a control group and five treatment groups. Seven of the ten beakers per group contain individual daphnids while the other three beakers per group contain multiple daphnids. Note that survival is plotted in these figures rather than mortality. For each group or concentration, the average survival in the seven individual beakers along with individual survival in the remaining three beakers is plotted. It is seen that in four of the six groups, individual survival exceeds multiple survival. There appears to be a downward trend in the survival rates associated with the beakers containing multiple daphnids. However for the individually housed daphnids there is no trend in survival rates. They remain at about 100 percent in the first five groups and then rapidly jump down to about 15 percent. Thus these plots suggest differences in mortality rates between the individually and multiply housed daphnids. There is also a suggestion of a possibly outlying response in group 5. This needs to be investigated further.

Figures II.11 to II.25 display trends in 21 day production vs group and vs concentration for the LeBlanc, Adams, Chapman, and Goulden tests. For the LeBlanc and Adams tests (with multiple daphnids per beaker) the fertility measure used is cumulative production per surviving adult per beaker. For the Chapman and Goulden tests (with individual daphnids per beaker) the fertility measure used is cumulative production for each individual daphnid that survived to the end of the test.

Figures II.11 to II.14 pertain to LeBlanc's tests with Compounds A and B. Figures II.11 and II.12 show an increase in productivity of the solvent control group in test A as compared to the water control group. There is then a decrease in production with increasing concentration. The highest treatment group has almost complete mortality (just one survivor of 80 daphnids that began the test) and no offspring. Figures II.13 and II.14 show first (among the treatment groups) an increase in production with increasing concentration and then a decrease. One of the values in group 3 appears to be a possible outlier. There does not appear to be heterogeneity of variance with increasing mean level.

Figures II.15 and II.16 pertain to Adams' test with selenium. We previously observed 100 percent mortality in groups 5 through 8 and we now see that there is no production in these groups either. (Virtually all the deaths in these groups occurred by day 4, well before any of the daphnids in the test were mature enough to produce offspring.) There appears to be a downward trend in average production with increasing concentrations in groups 1 to 4. There is also a suggestion of increase in variability.

Figures II.17 to II.20 pertain to Chapman's test with beryllium. Figure II.17 displays total young produced by each adult surviving to the end of the test. The water control and solvent control groups appear to be comparable and higher on average than the six treatment groups. There is no suggestion of heterogeneity of variance or of outlying values. There is no apparent trend in production levels among the six treatment groups. Figures II.18 and II.19 display average total production for surviving

adults versus group number and versus $\log_{10}$ (concentration). In both plots we see that the control groups' production levels are much higher than the treatment group production levels, the solvent control group has higher average production level than the water control group, and the treatment group average production levels show no trend with increasing concentration. Figure II.20 shows the standard deviation of 21 day production levels of survivors plotted against average production levels. There is an increasing relationship between standard deviation and average level.

Figures II.21 to II.25 pertain to Goulden's test with isophorone. Figure II.21 displays individual 21 day production for all daphnids, both survivors and nonsurvivors. Recall that production data were recorded only on the seven individually housed daphnids per group. We see a definite quadratic trend in the plot, first increasing and then decreasing. Figures II.22, II.23, and II.24 display 21 day production for all individually housed daphnids that survived to the end of the tests. Figure II.22 is nearly identical in appearance to Figure II.21 since there was just one death among individually housed daphnids in groups 1 to 5 and in group 6 there was no production among any of the seven daphnids, whether or not they survived. Figure II.25 displays standard deviation of production versus average production for the surviving daphnids. No trend is evident. Note that group 6 is not represented in this plot since there was just one surviving daphnid and so the standard deviation was not calculated.

Figures II.26 to II.34 display trends in average 21 day lengths and in variation within beakers and among beakers within groups for the LeBlanc and Chapman tests. Adams and Goulden do not report length data.

Figures II.26, II.27, and II.28 pertain to LeBlanc's test with Compound A. Figure II.26 shows a higher average length among survivors in the solvent control group than in the water control group. No trend in either average length or variability of length is evident in groups 3 to 6. The point in group 7 corresponds to just a single daphnid, the only survivor in that group, and is relatively stunted. Figure II.27 displays the within beaker standard deviations of length versus the within beaker average lengths. A negative trend can be seen, indicating that variability decreases with increasing length! This is opposite to what occurs with most physical phenomena. Figure II.28 displays standard deviations among beaker averages within groups versus means of beaker averages. A clear, strong negative trend is evident; that is, decreasing variability with increasing length.

Figures II.29, II.30, and II.31 pertain to LeBlanc's test with Compound B. Figure II.29 shows a lower average length among survivors in the solvent control group than in the water control group. This is opposite to what was observed for the Compound A test. This suggests that no universal relation is appropriate. There is little trend in either average length or in variability of length in groups 3 to 6. The lengths in group 7 are clearly stunted. Figure II.30 displays the within beaker standard deviations of length versus within beaker averages. There is little trend in the plot. If any exists, it is slightly negative. Figure II.31 displays

the standard deviation among beaker averages within groups versus mean of beaker averages. A negative trend is evident, although not as strong as for Compound A. The point to the far left of the plot corresponds to group 7, which is not typical of the other groups. However a negative trend exists even without this point.

Figures II.32, II.33, and II.34 pertain to Chapman's test with beryllium. Figures II.32 and II.33 show about the same average lengths in the solvent control and in the water control groups and a decreasing trend in average length with increasing concentration. The variability among lengths decreases as the average among lengths increases. This is the same phenomenon that we saw with Compound A. Figure II.34 displays the standard deviation among lengths within groups versus average legnth. There is a clear negative trend.

We thus conclude that the variability of daphnid lengths decreases as average length increases. This type of phenomenon would be compatible with a biological upper limit on daphnid length that the adult daphnids are approaching. The distribution of daphnid lengths would then be skewed to the left.

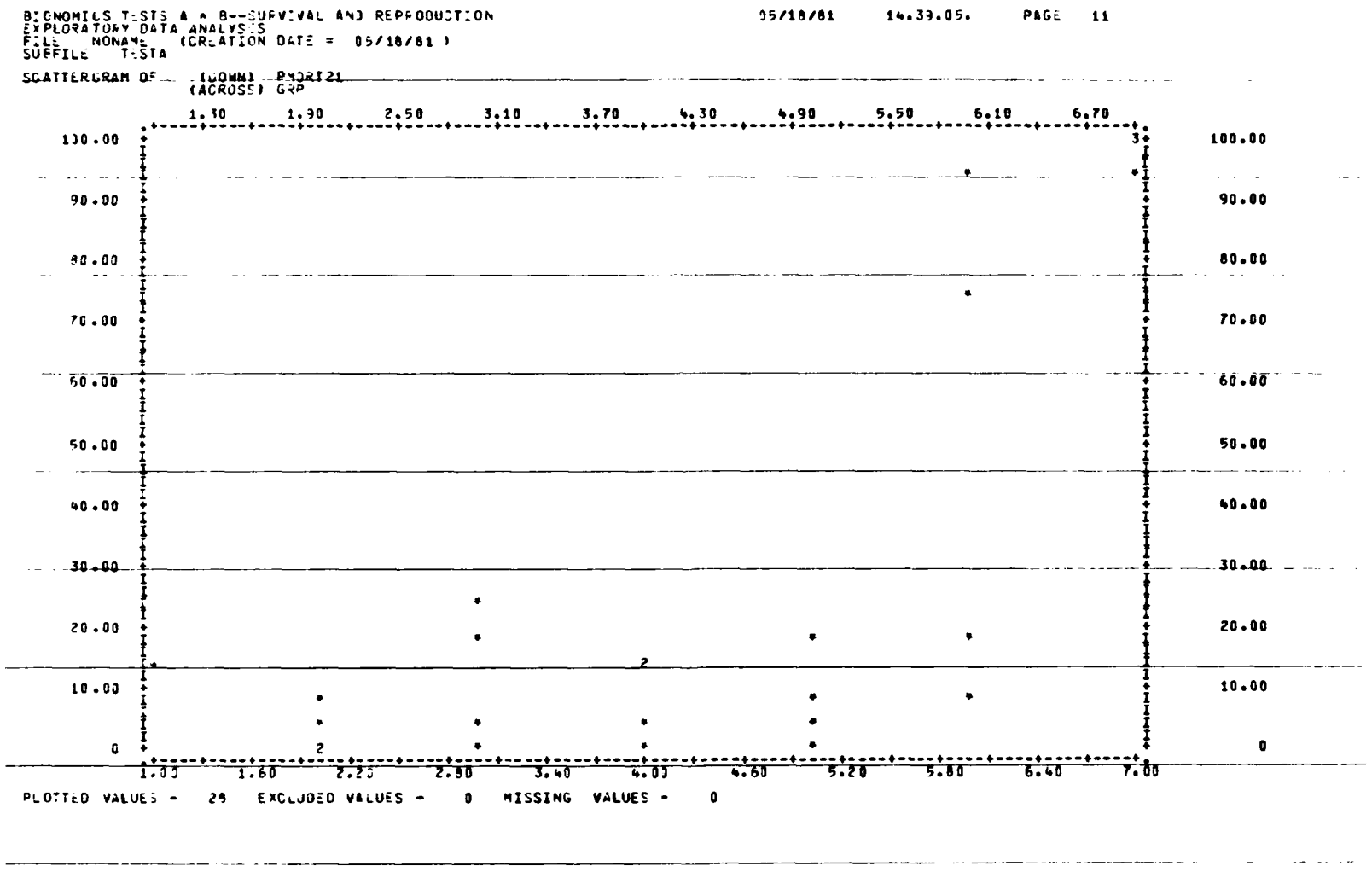


Figure II.1. LeBlanc - Test A. 21 day mortality by treatment group

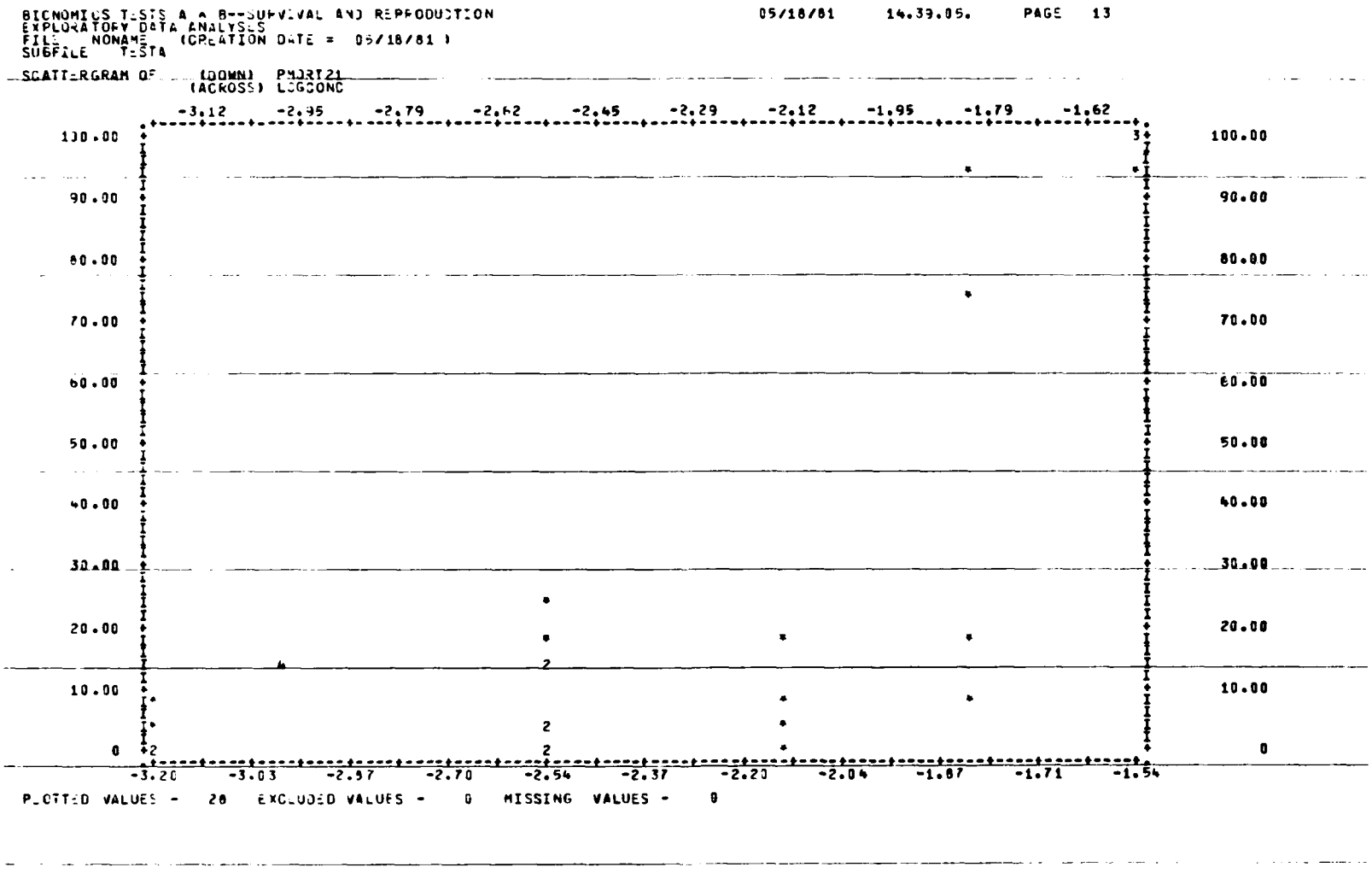


Figure II.2. LeBlanc - Test A. 21 day mortality by $\log_{10}$ (concentration)

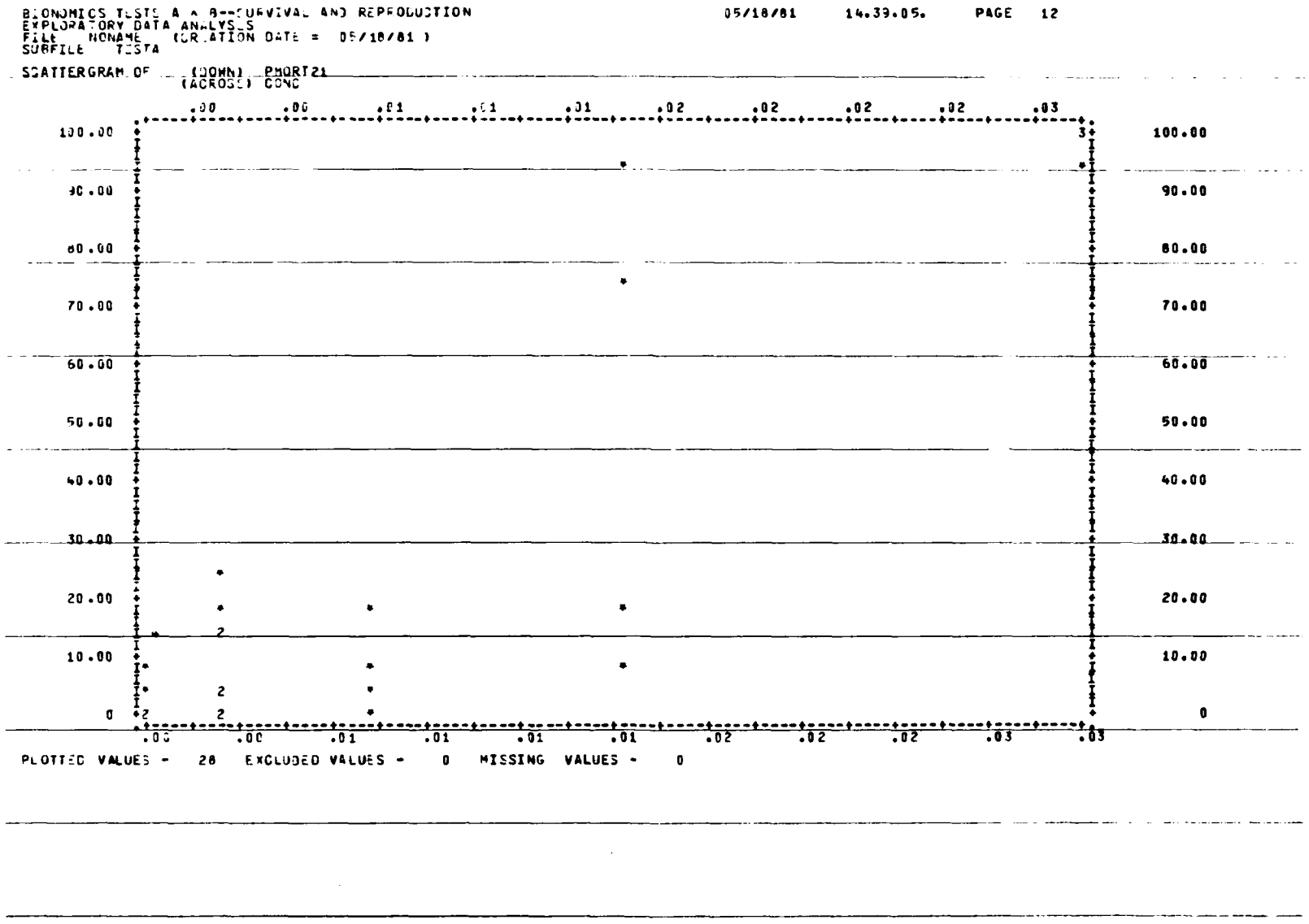


Figure II.2a. LeBlanc - Test A. 21 day mortality by concentration

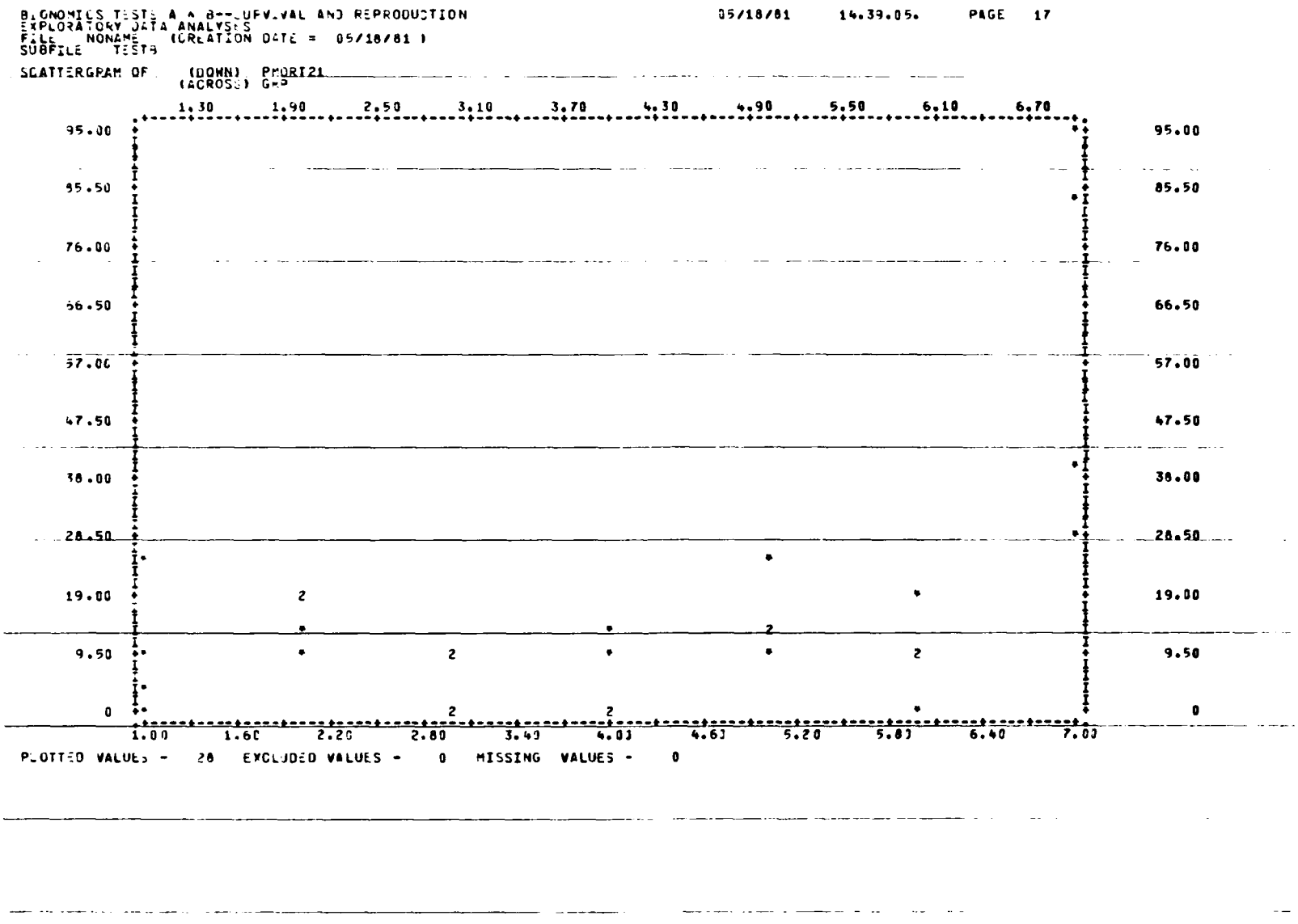


Figure II.3. LeBlanc - Test A. 21 day mortality by treatment group

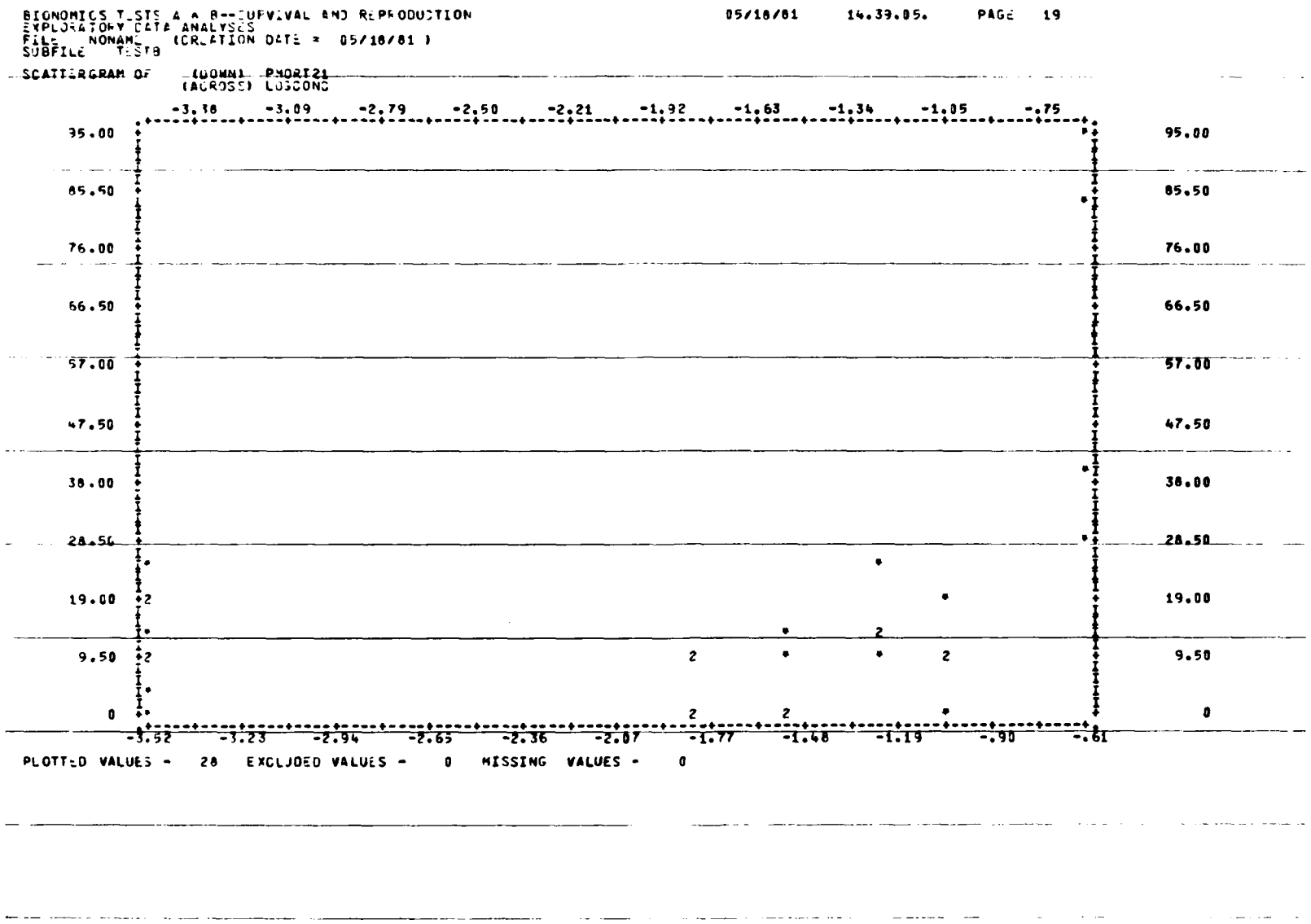


Figure II.4. LeBlanc - Test B. 21 day mortality by $\log_{10}$ (concentration)

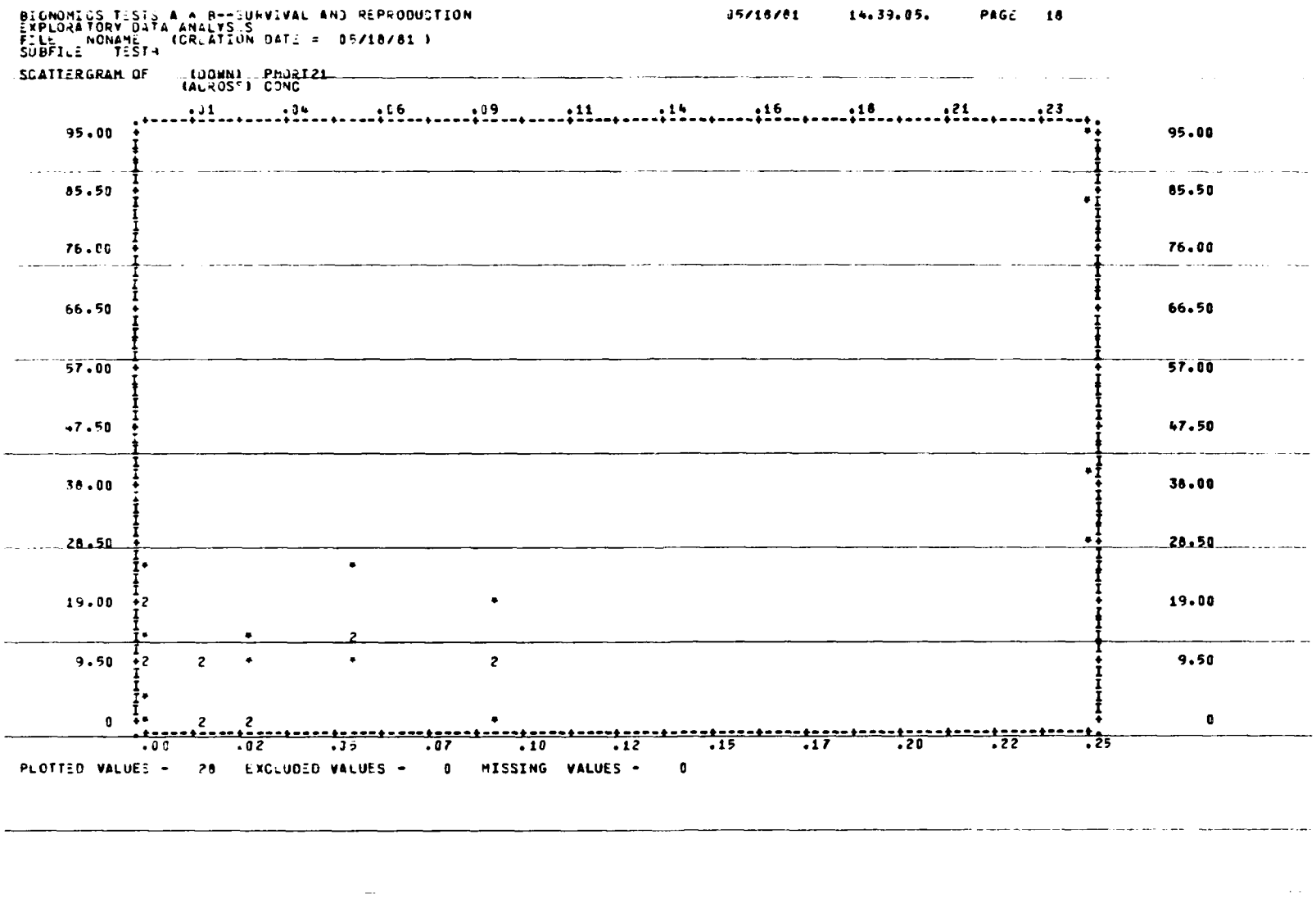


Figure II.4a. LeBlanc - Test B. 21 day mortality by concentration

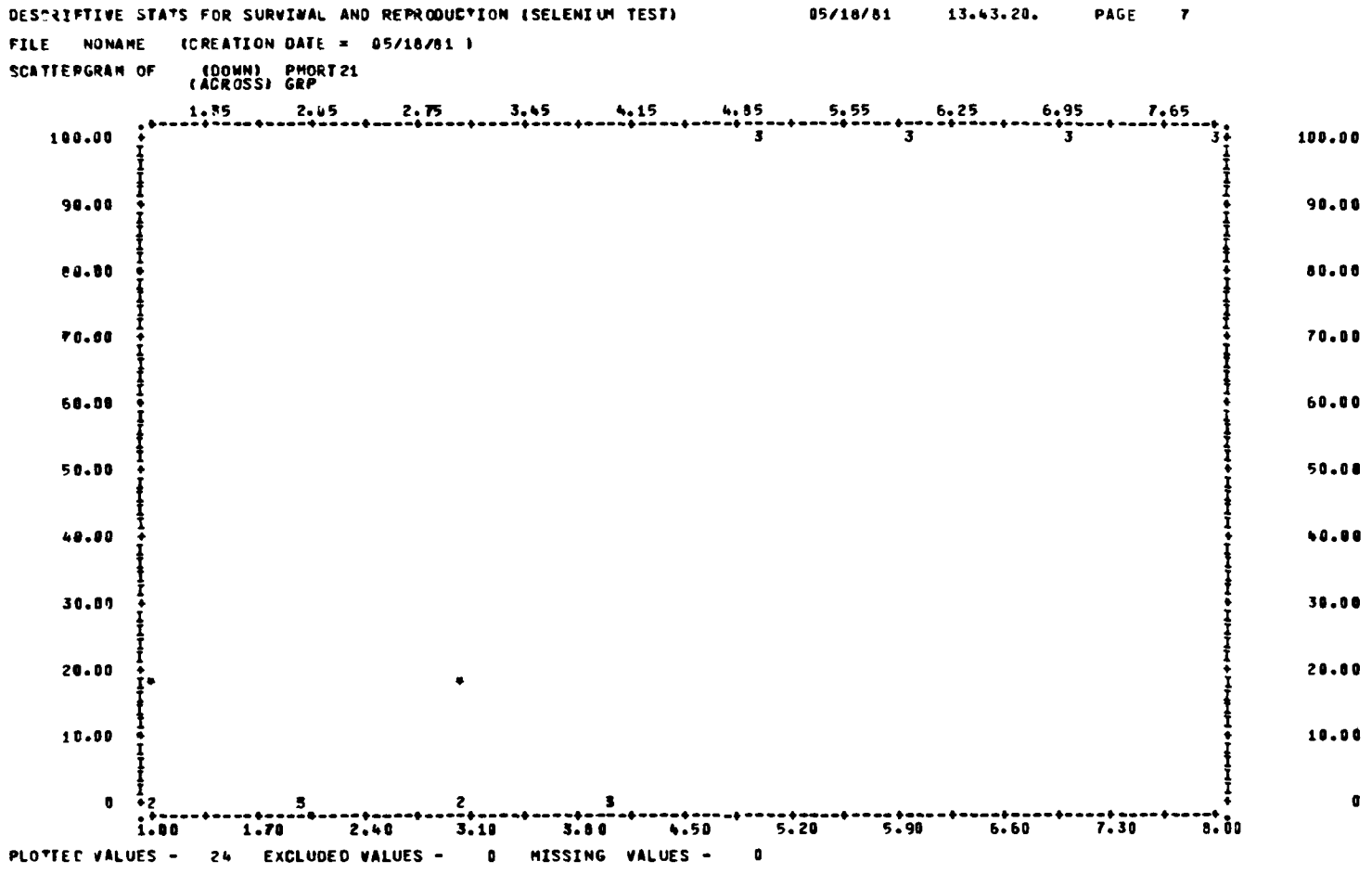


Figure II.5. Adams - Selenium. 21 day mortality by group

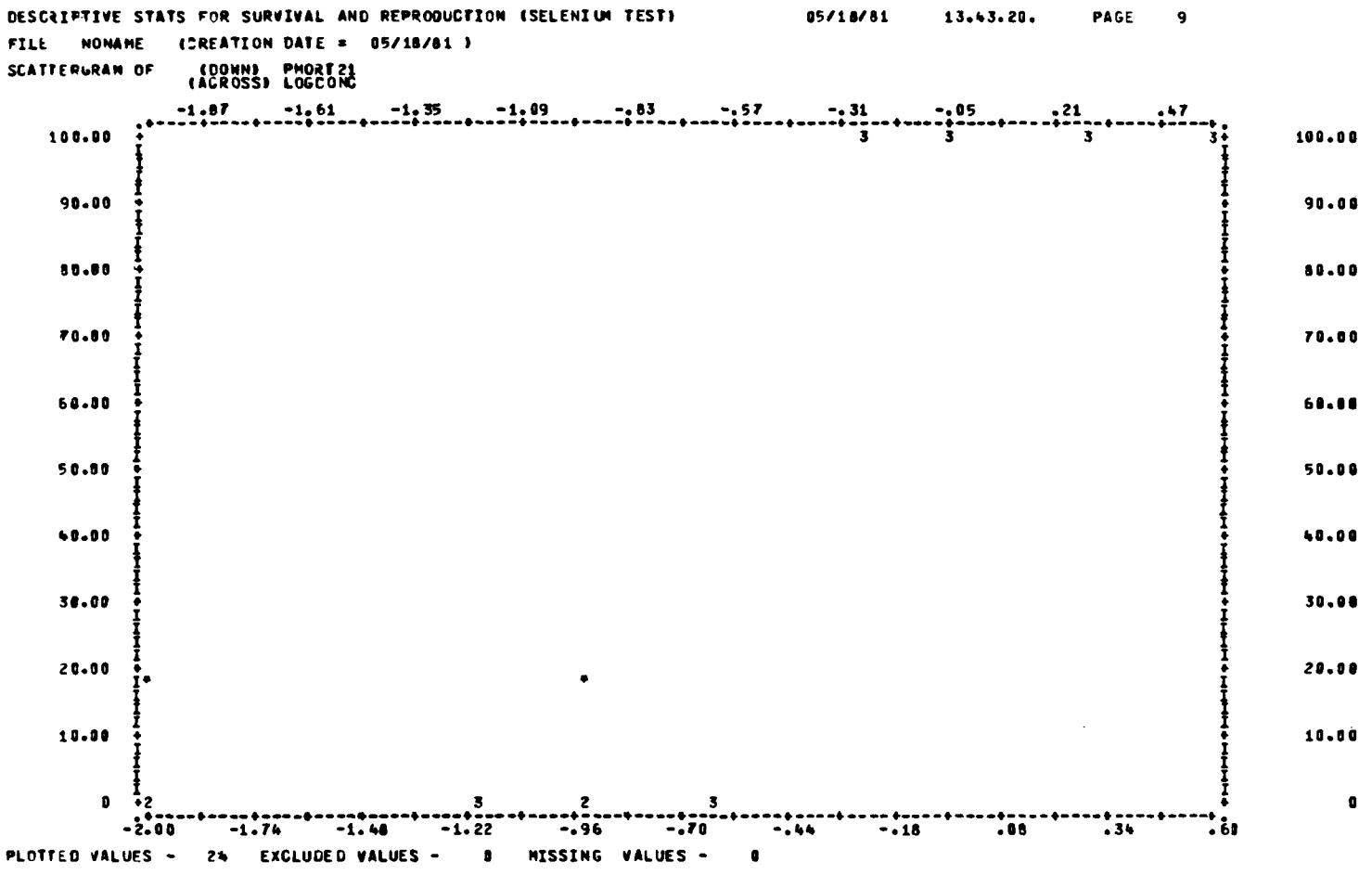


Figure II.6. Adams - Selenium. 21 day mortality by $\log_{10}$ (concentration)

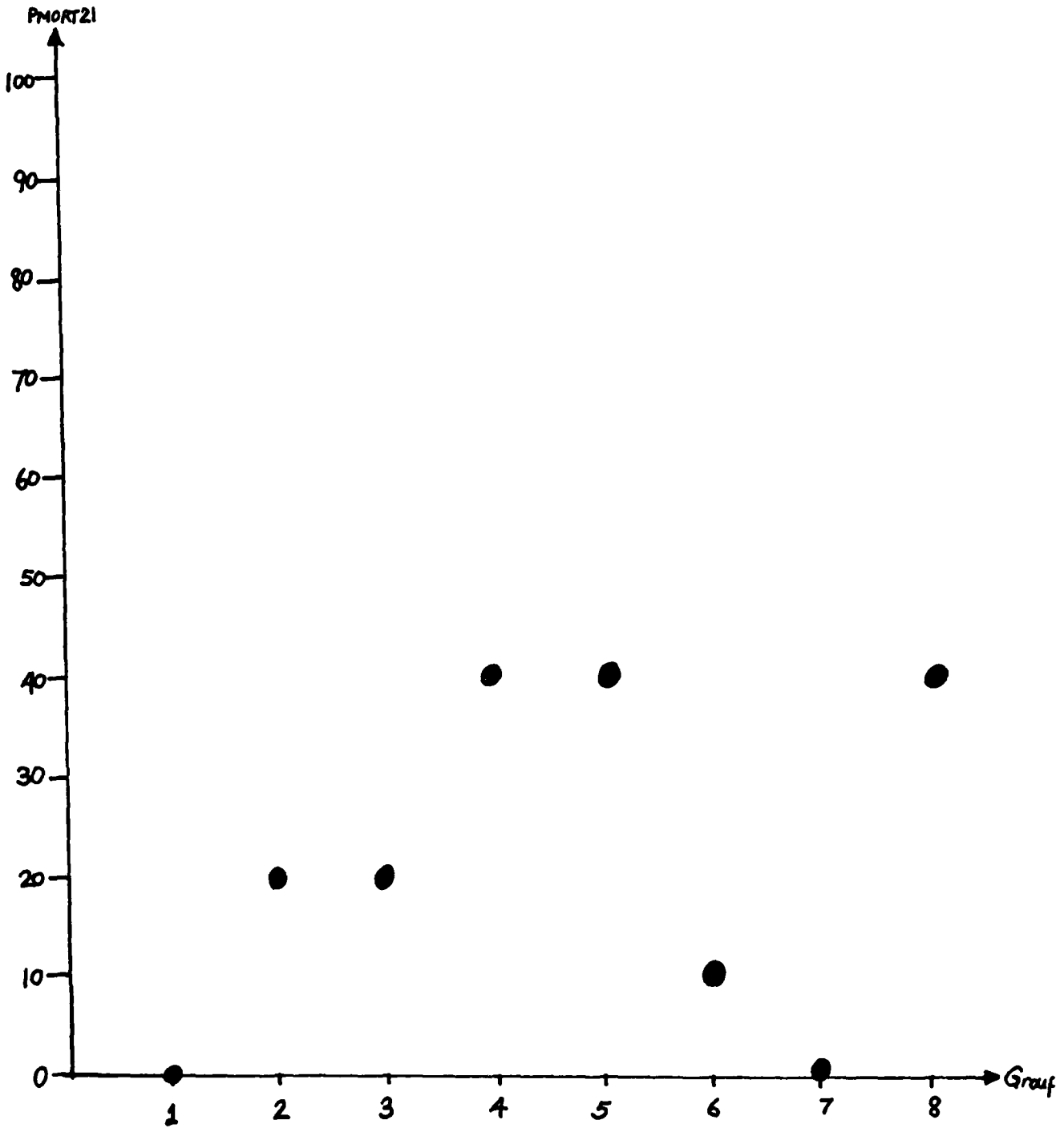


Figure II.7. Chapman - Beryllium. 21 day morality by group

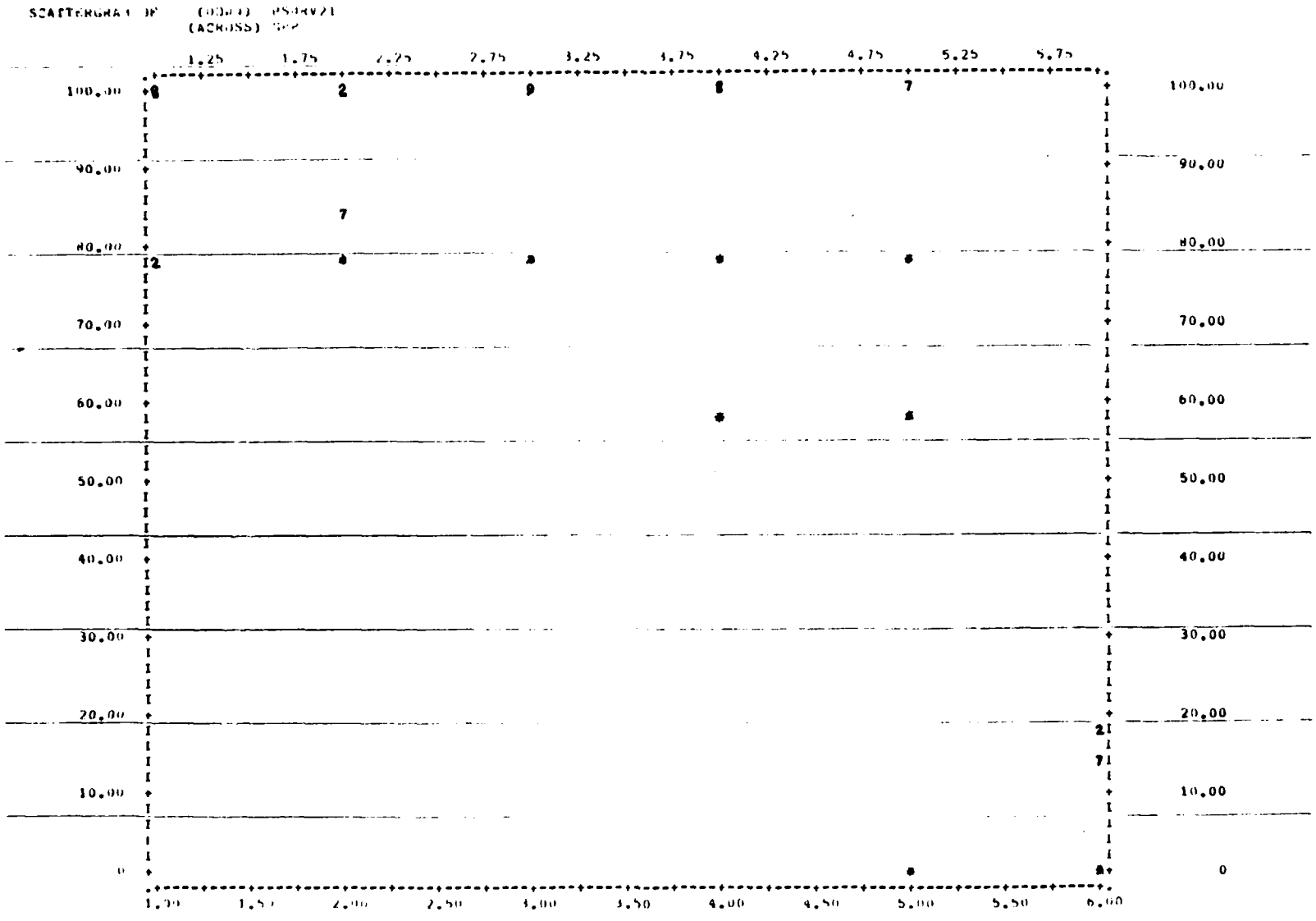


Figure II.8. Goulden - Isophorone. 21 day survival by group

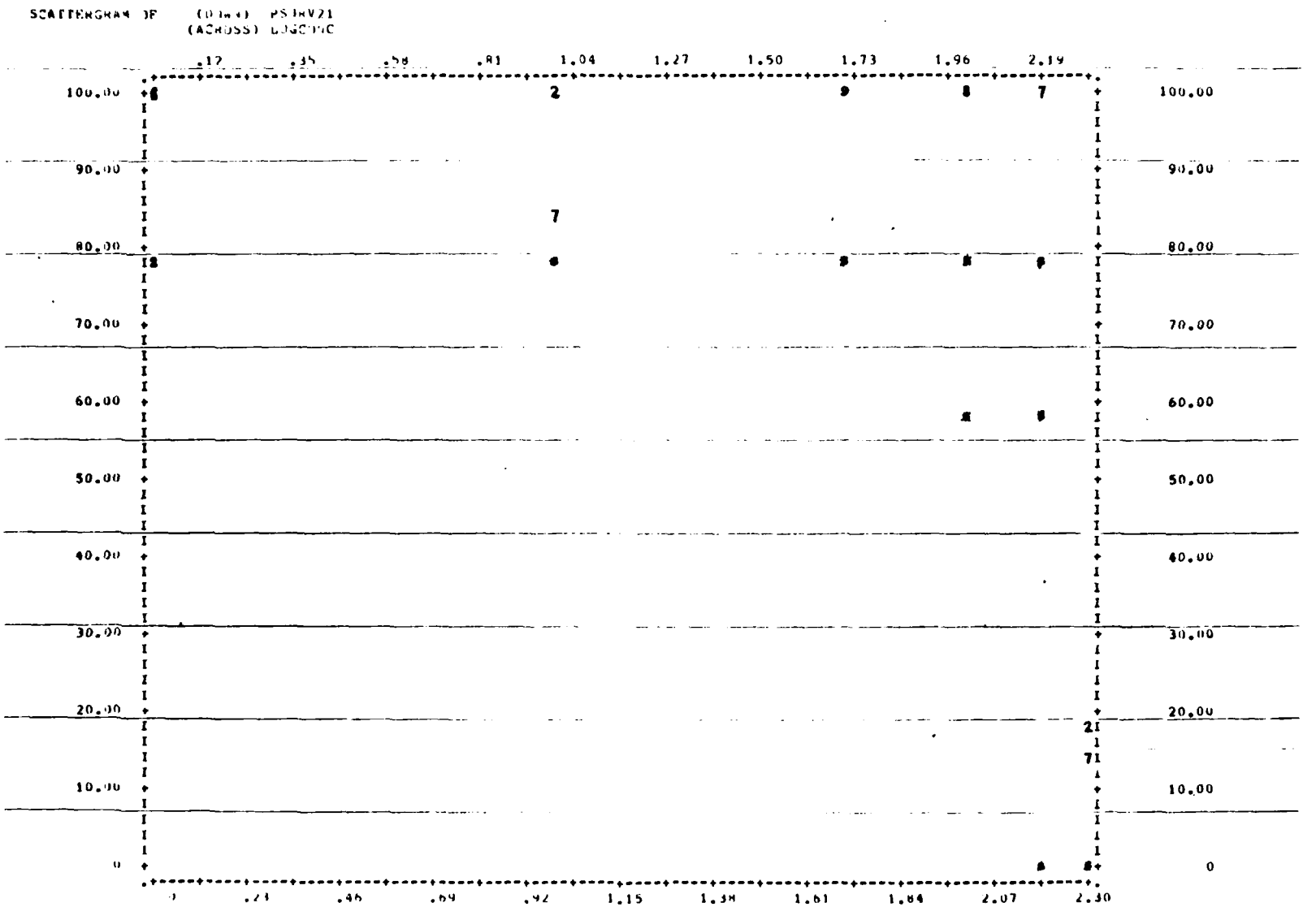


Figure II.9. Goulden - Isophorone. 21 day survival by $\log_{10}$ (1 + concentration)

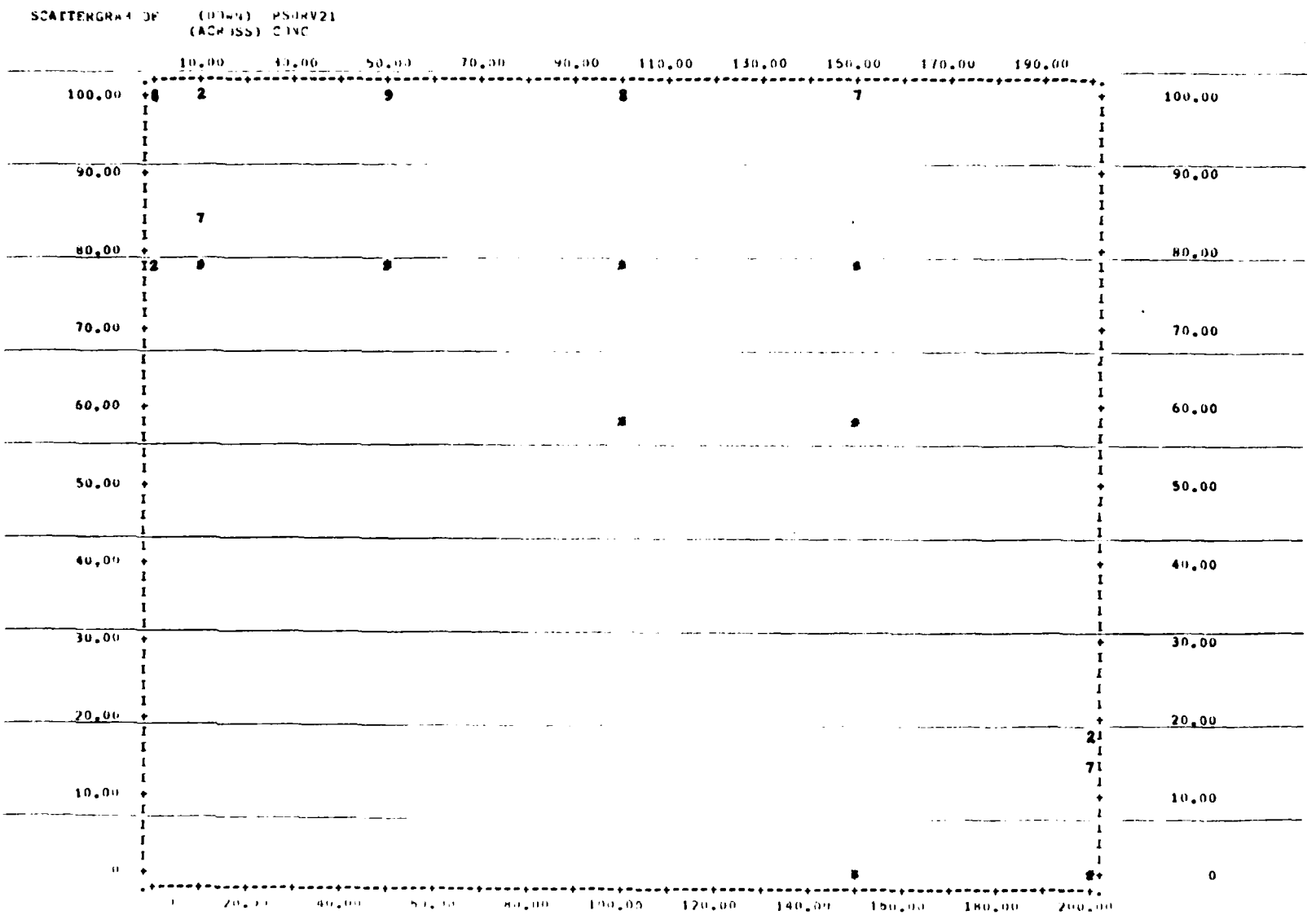


Figure II.10. Goulden - Isophorone. 21 day survival by concentration

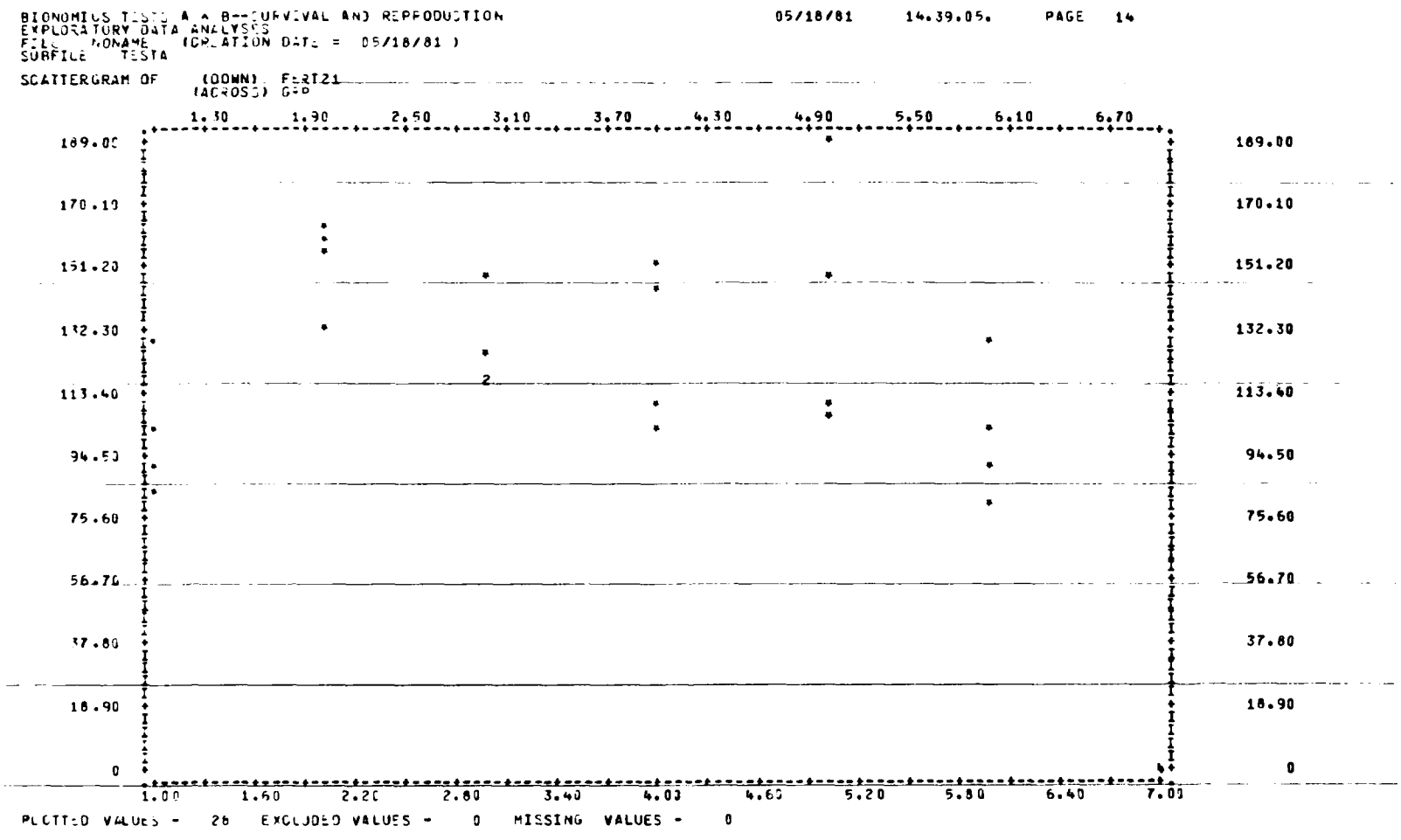


Figure II.11. LeBlanc - Test A. 21 day cumulative fertility per surviving adult by group

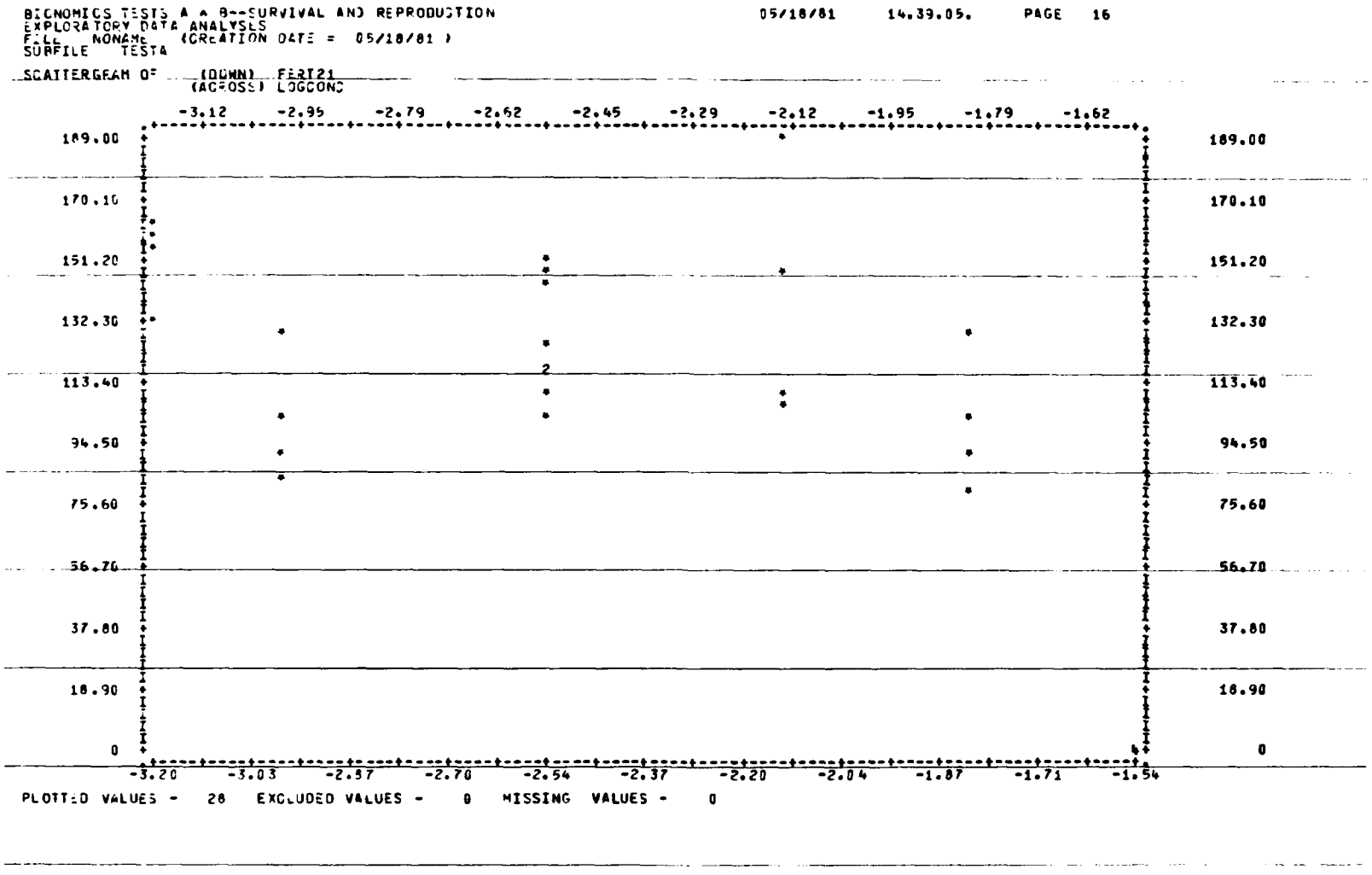


Figure II.12. LeBlanc - Test A. 21 day cumulative fertility per surviving adult by $\log_{10}$ (concentration)

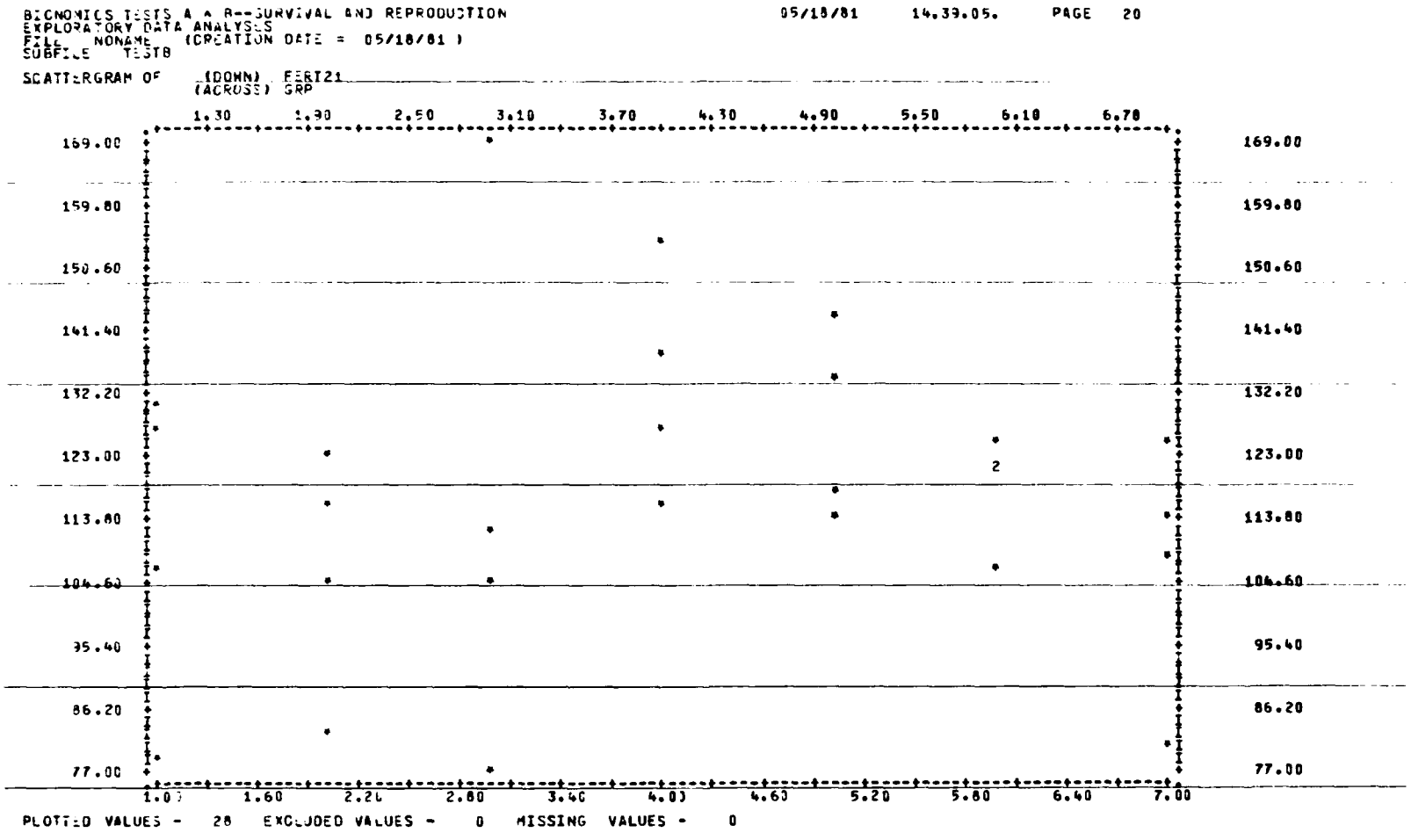


Figure II.13. LeBlanc - Test B. 21 day cumulative fertility per surviving adult by group

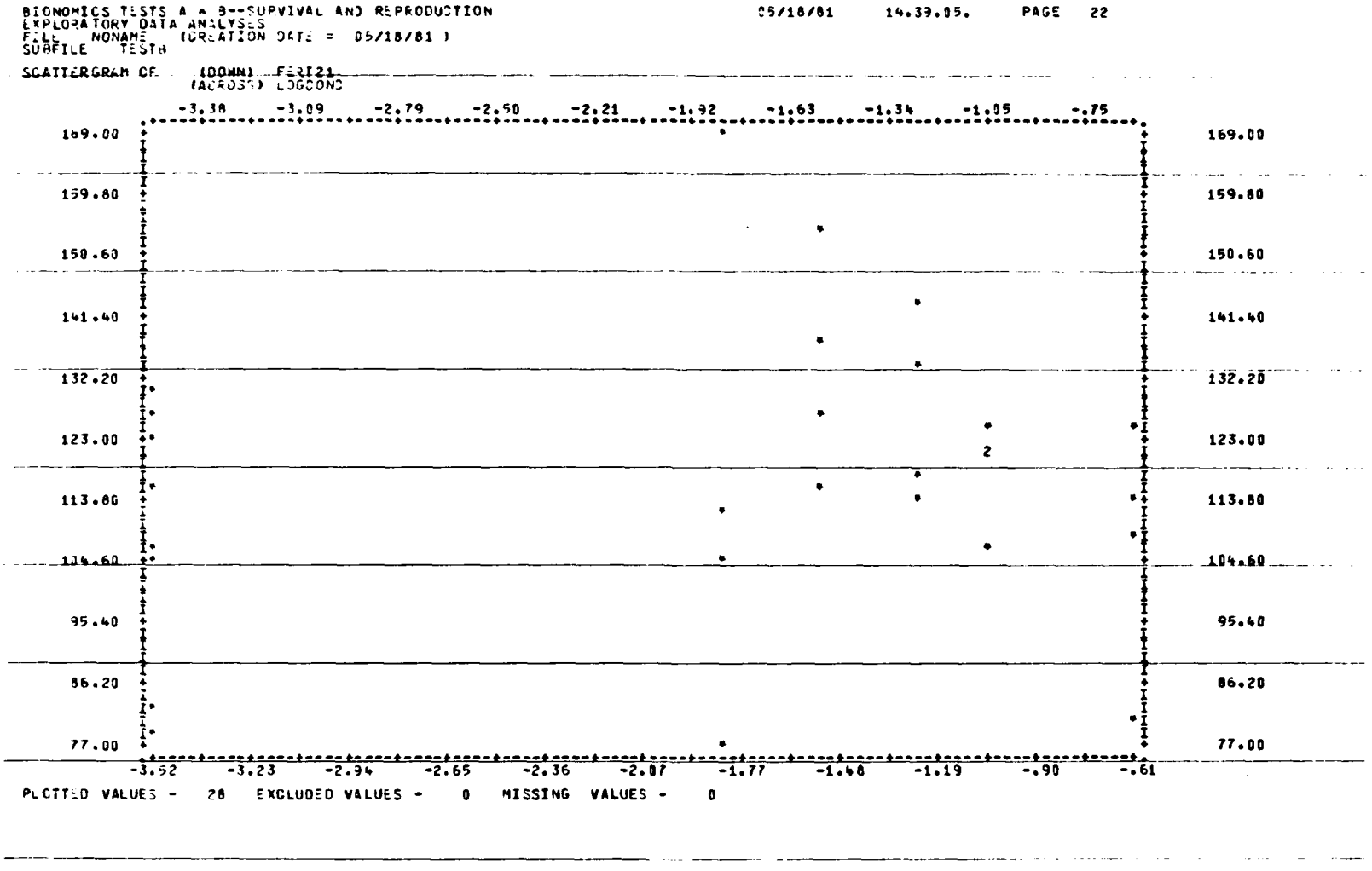


Figure II.14. LeBlanc - Test B. 21 day cumulative fertility per surviving adult by $\log_{10}$ (concentration)

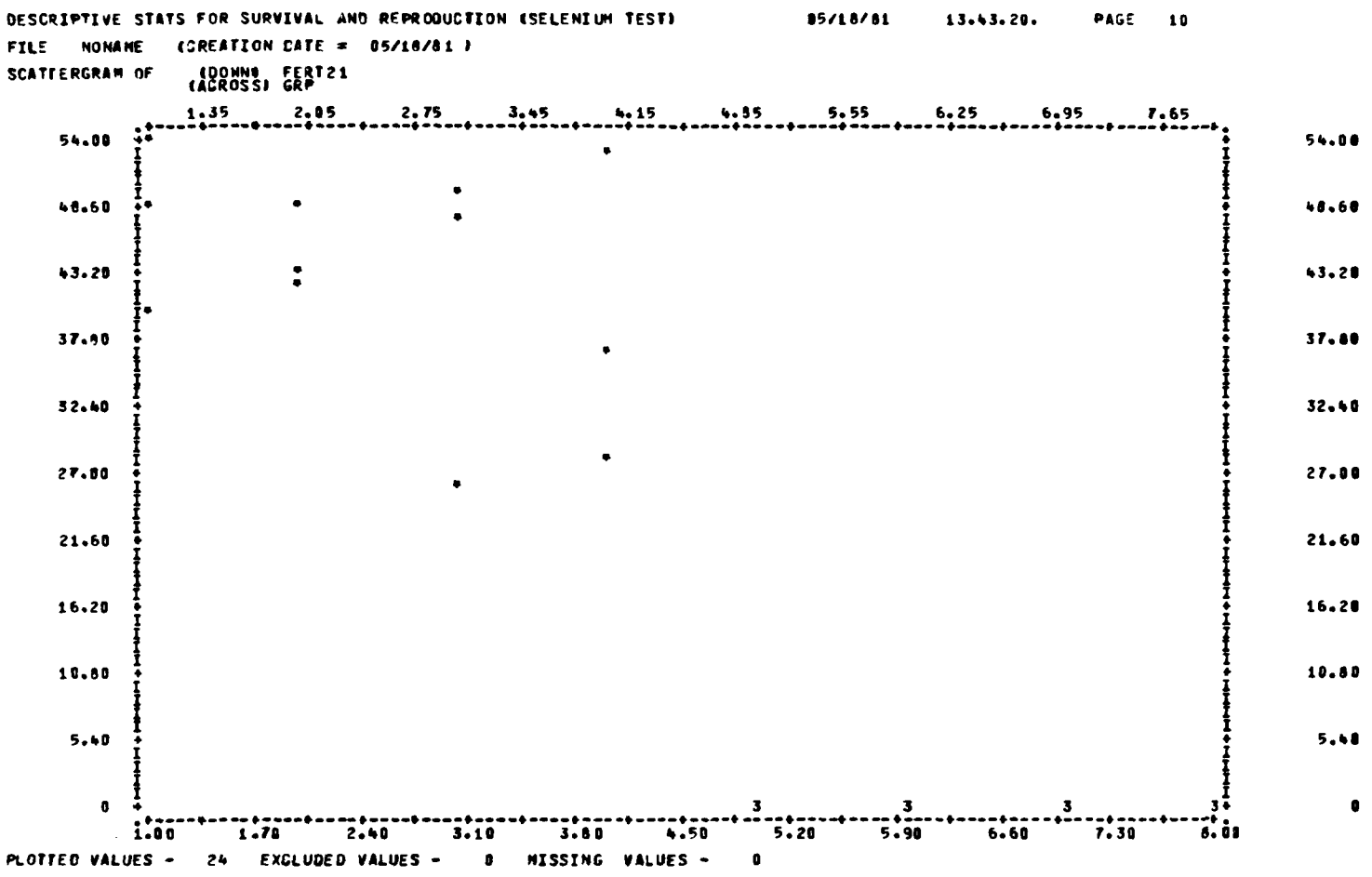


Figure II.15. Adams - Selenium. 21 day cumulative fertility per surviving adult by group

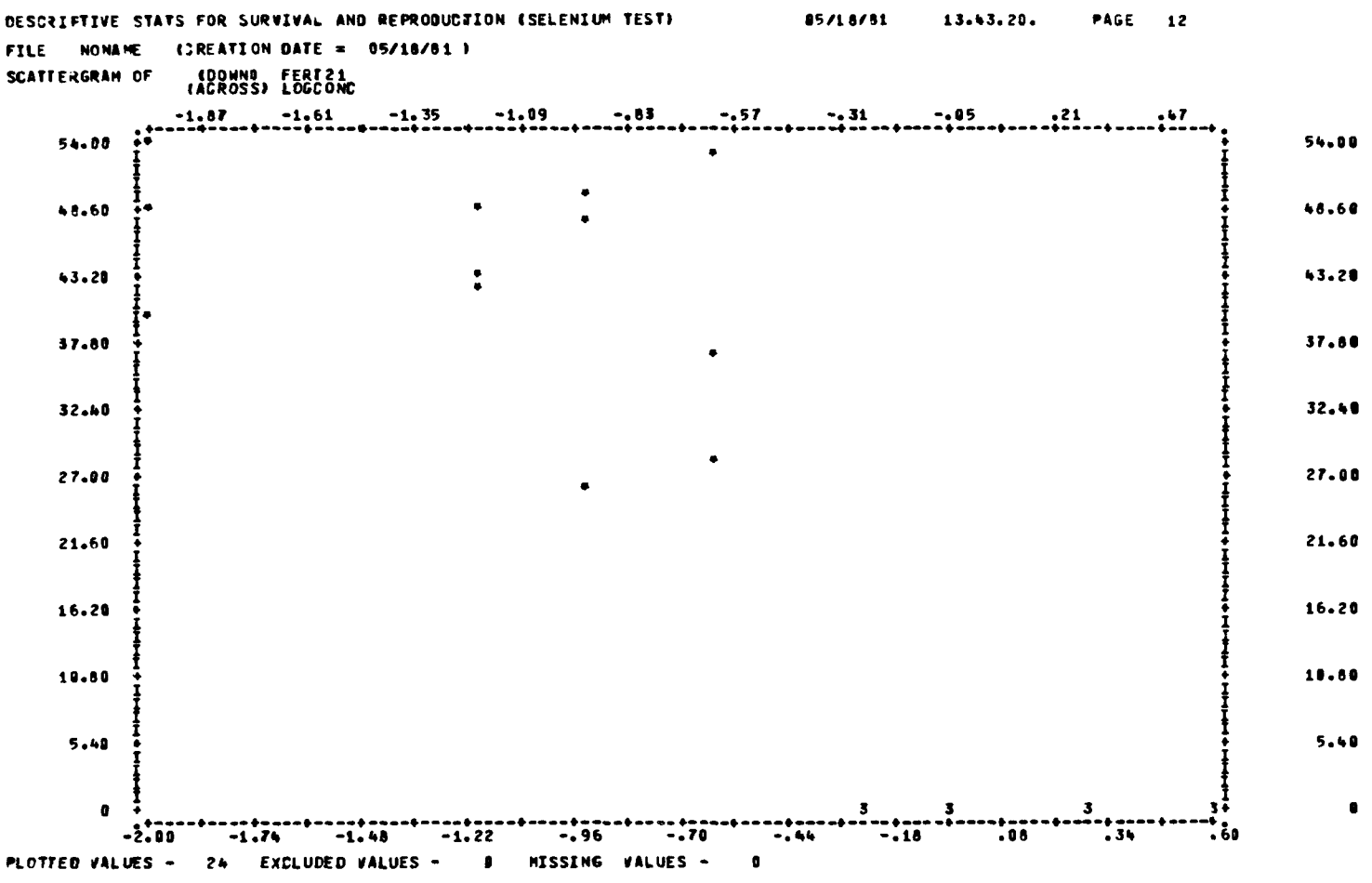


Figure II.16. Adams - Selenium. 21 day cumulative fertility per surviving adult by $\log_{10}$ (concentration)

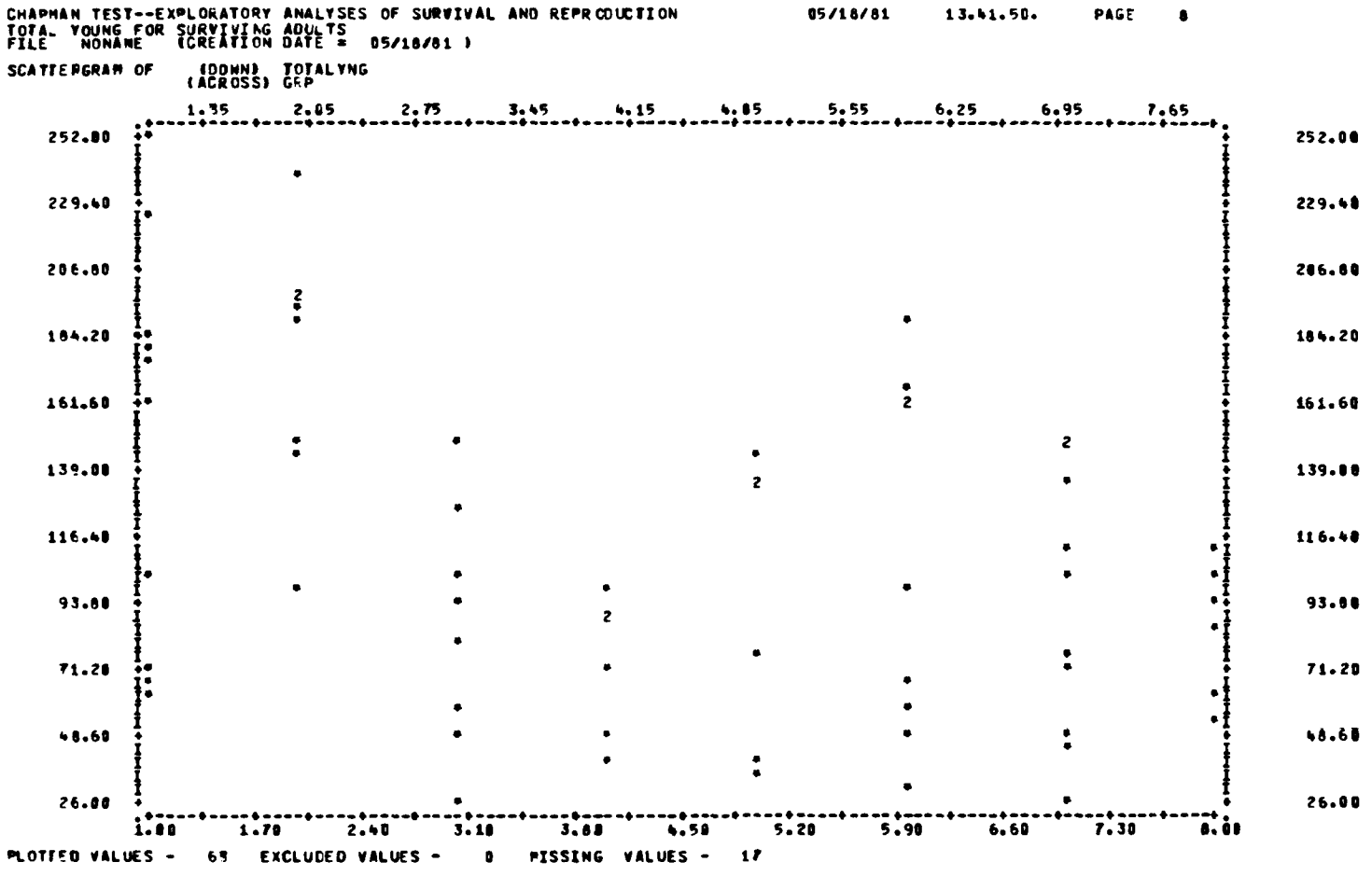


Figure II.17. Chapman - Beryllium. Individual 21 day production for surviving adults by group

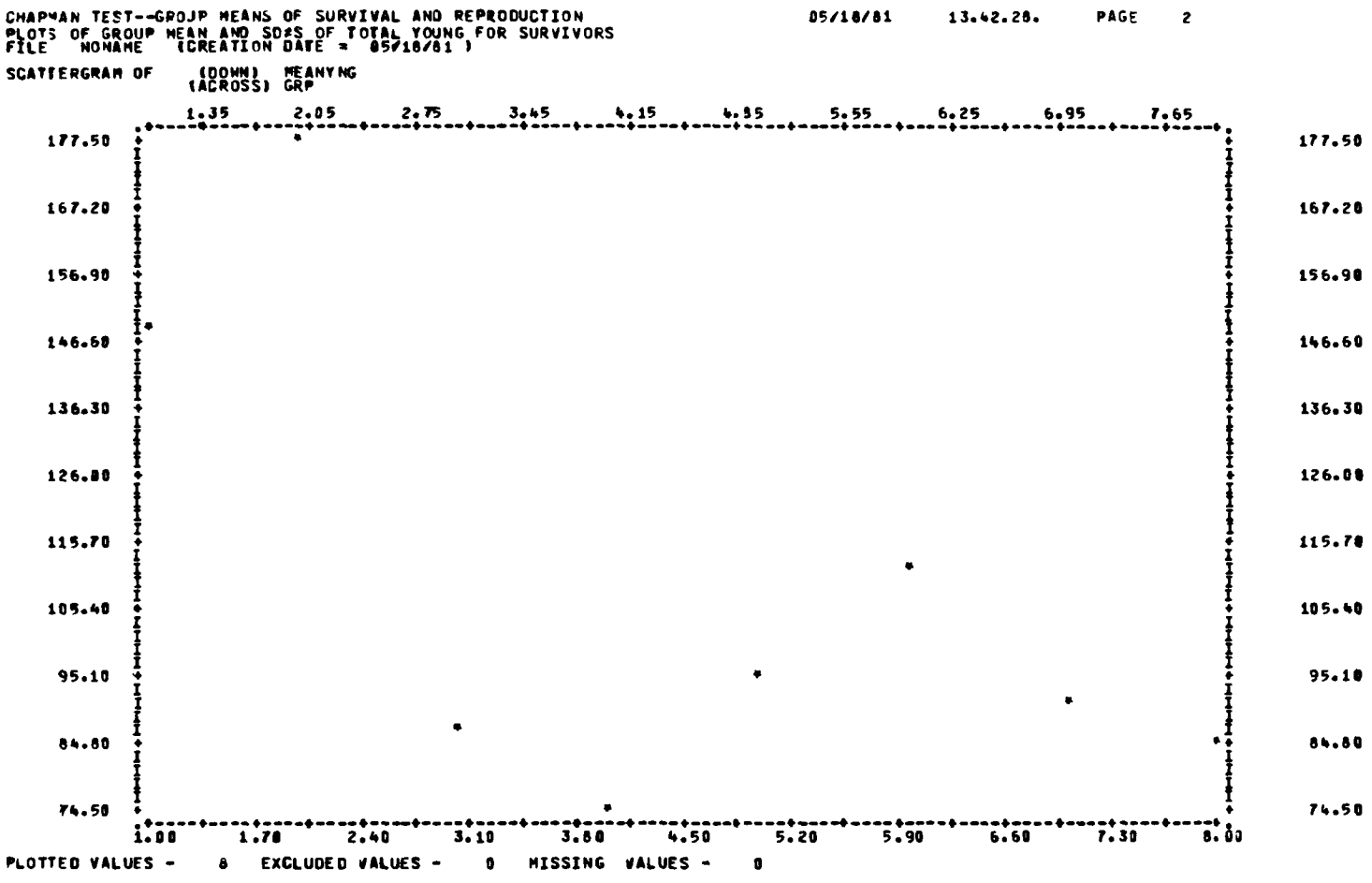


Figure II.18. Chapman - Beryllium. Average 21 day production for surviving adults by group

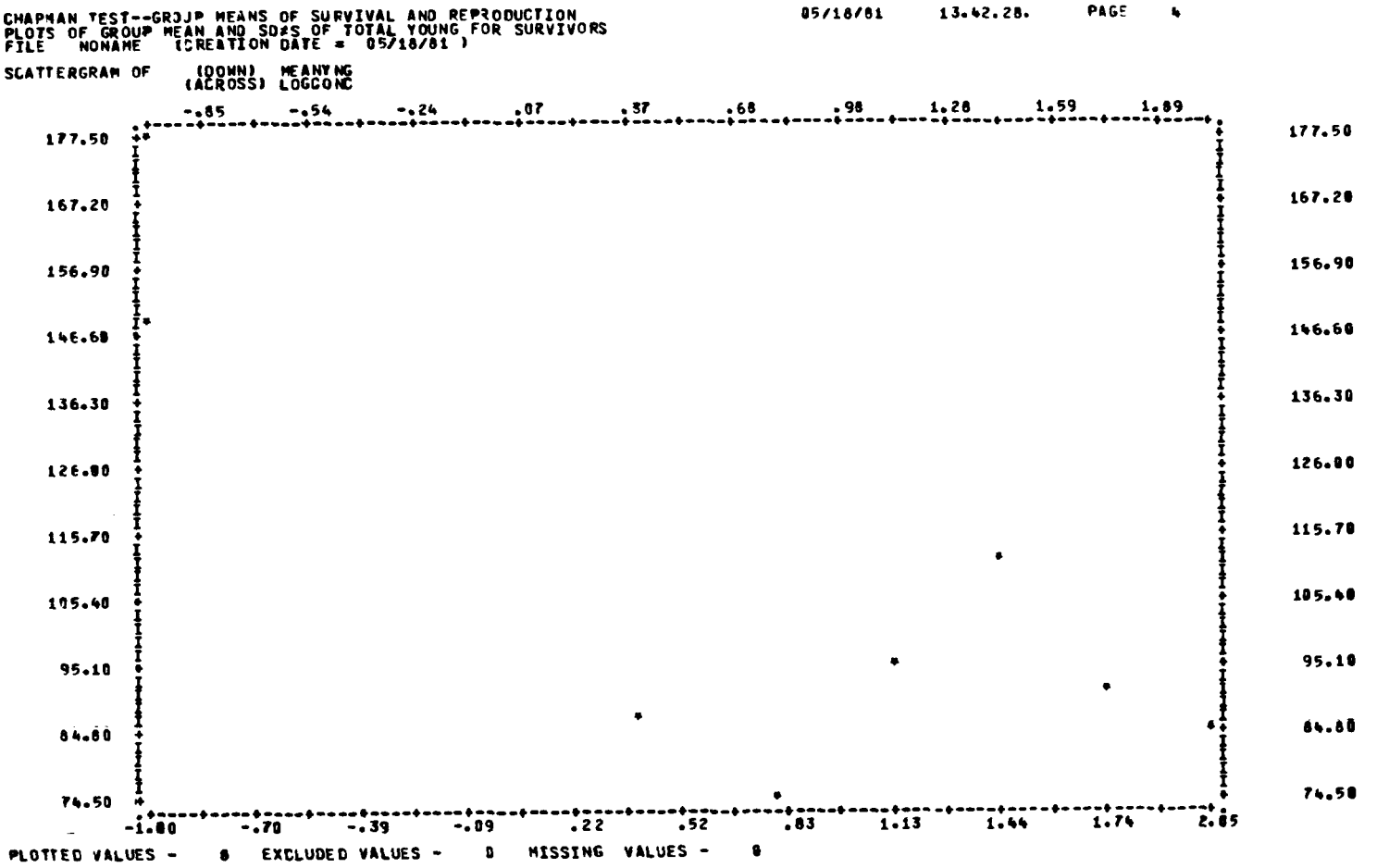
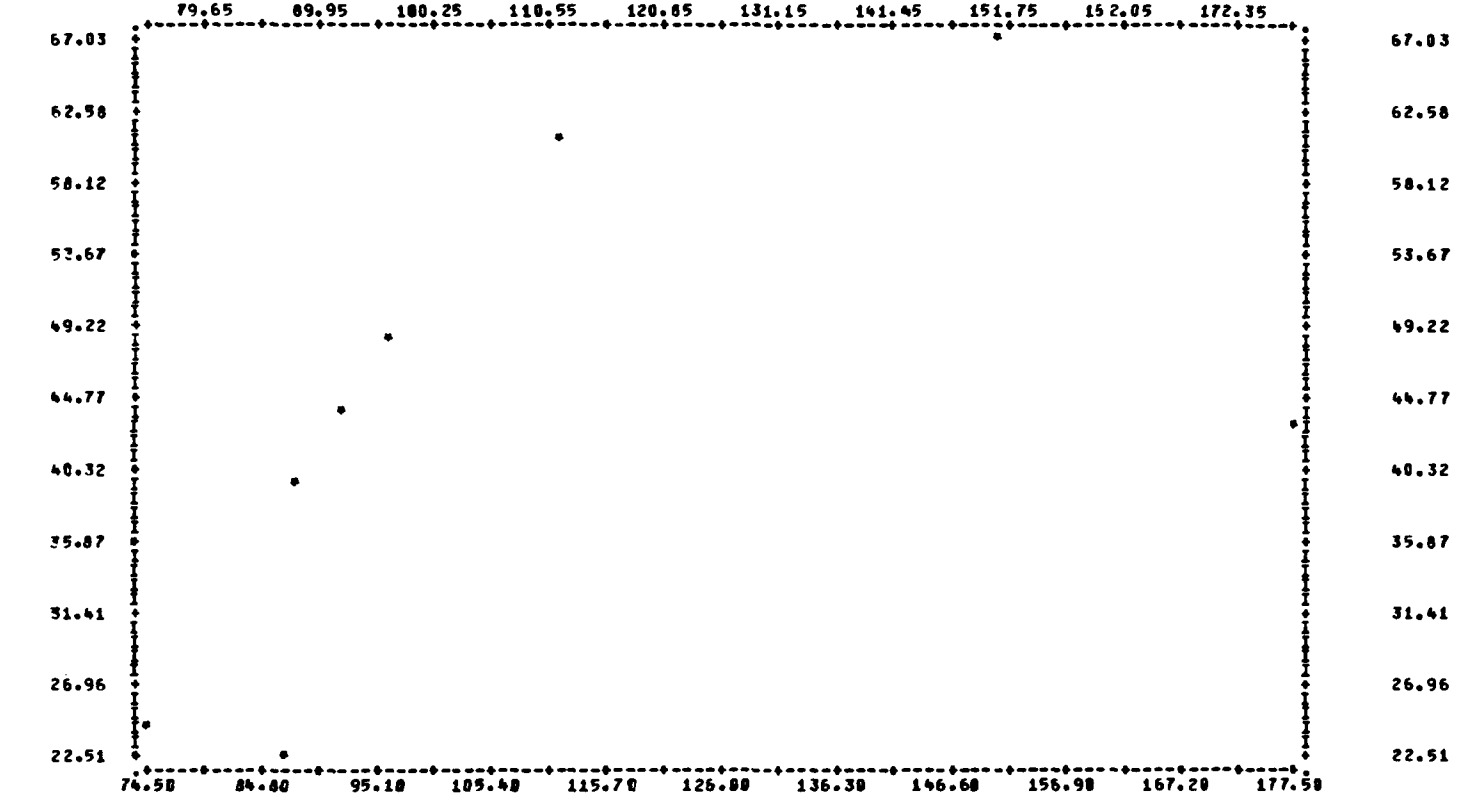


Figure II.19. Chapman - Beryllium. Average 21 day production per surviving adult by $\log_{10}$ (concentration)

CHAPMAN TEST--GROUP MEANS OF SURVIVAL AND REPRODUCTION 05/18/81 13.42.28. PAGE 9
REPRODUCTION--EXAMINING HETEROGENEITY OF GROUP VARIANCES
FILE NONAME (CREATION DATE = 05/18/81)

SCATTERGRAM OF (DOWN) STDDEV
(ACROSS) MEANYNG



STATISTICS..

CORRELATION (R)-	.54334	R SQUARED -	.29521	SIGNIFICANCE R -	.08139
STD ERR OF EST -	14.16708	INTERCEPT (A) -	18.04560	STD ERROR OF A -	17.12520
SIGNIFICANCE A -	.16629	SLOPE (B) -	.23674	STD ERROR OF B -	.14933
SIGNIFICANCE B -	.08199				

PLOTTED VALUES - 8 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure II.20. Chapman - Beryllium. Standard deviation of 21 day production for surviving adults by average production

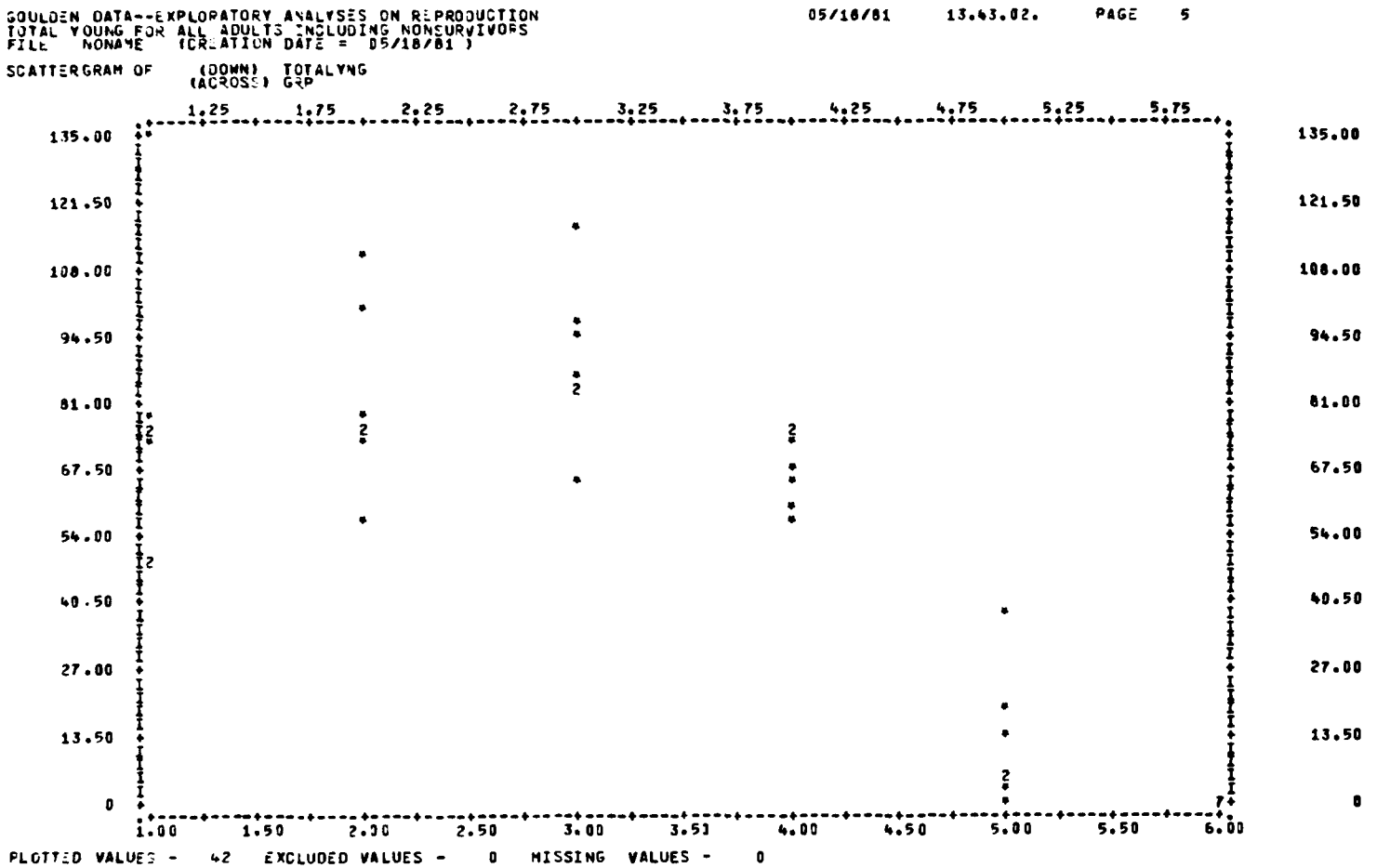


Figure II.21. Goulden - Isophorone. Individual 21 day production for both surviving and nonsurviving adults by group

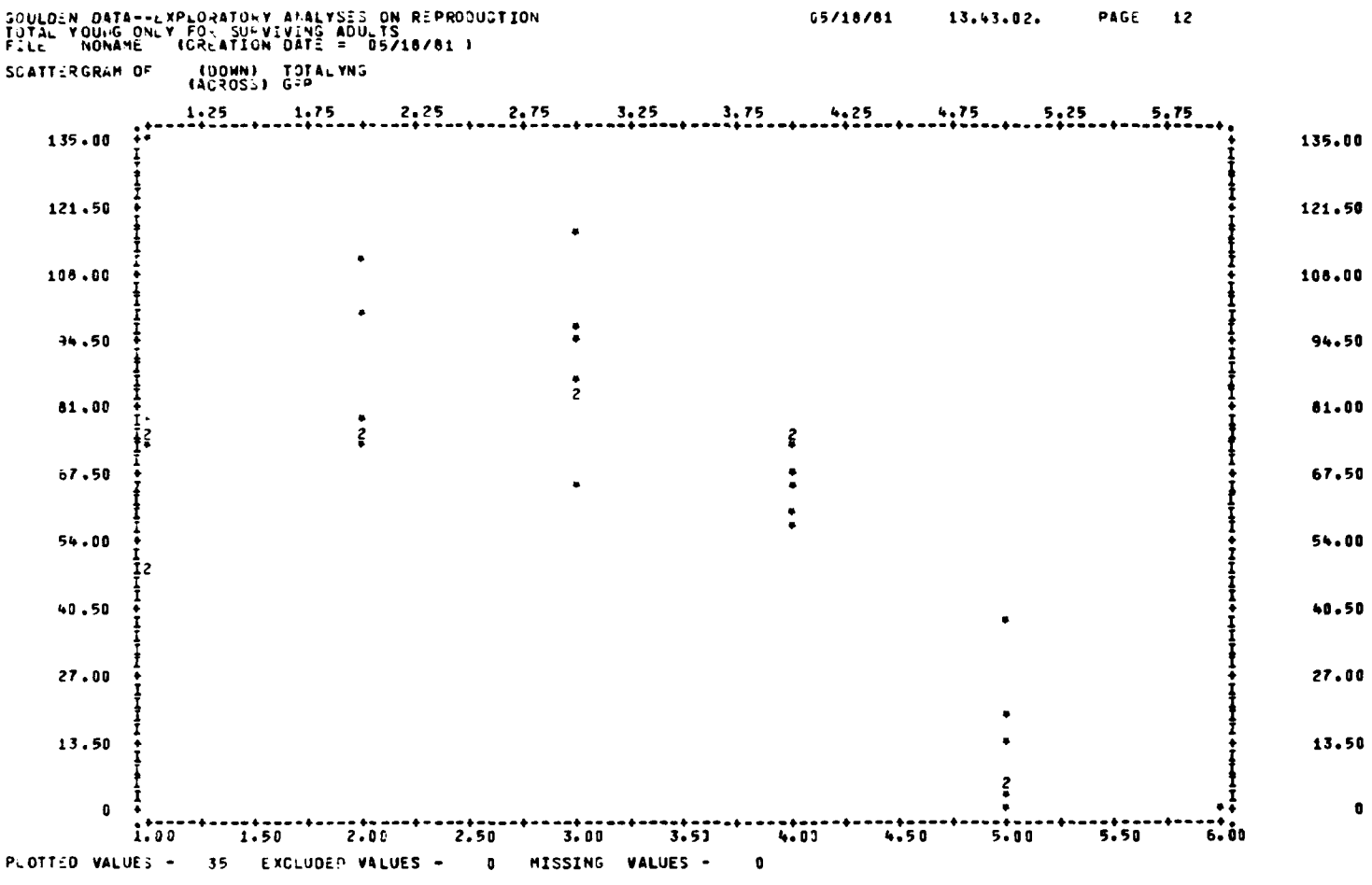


Figure II.22. Goulden - Isophorone. Individual 21 day production for surviving adults by group

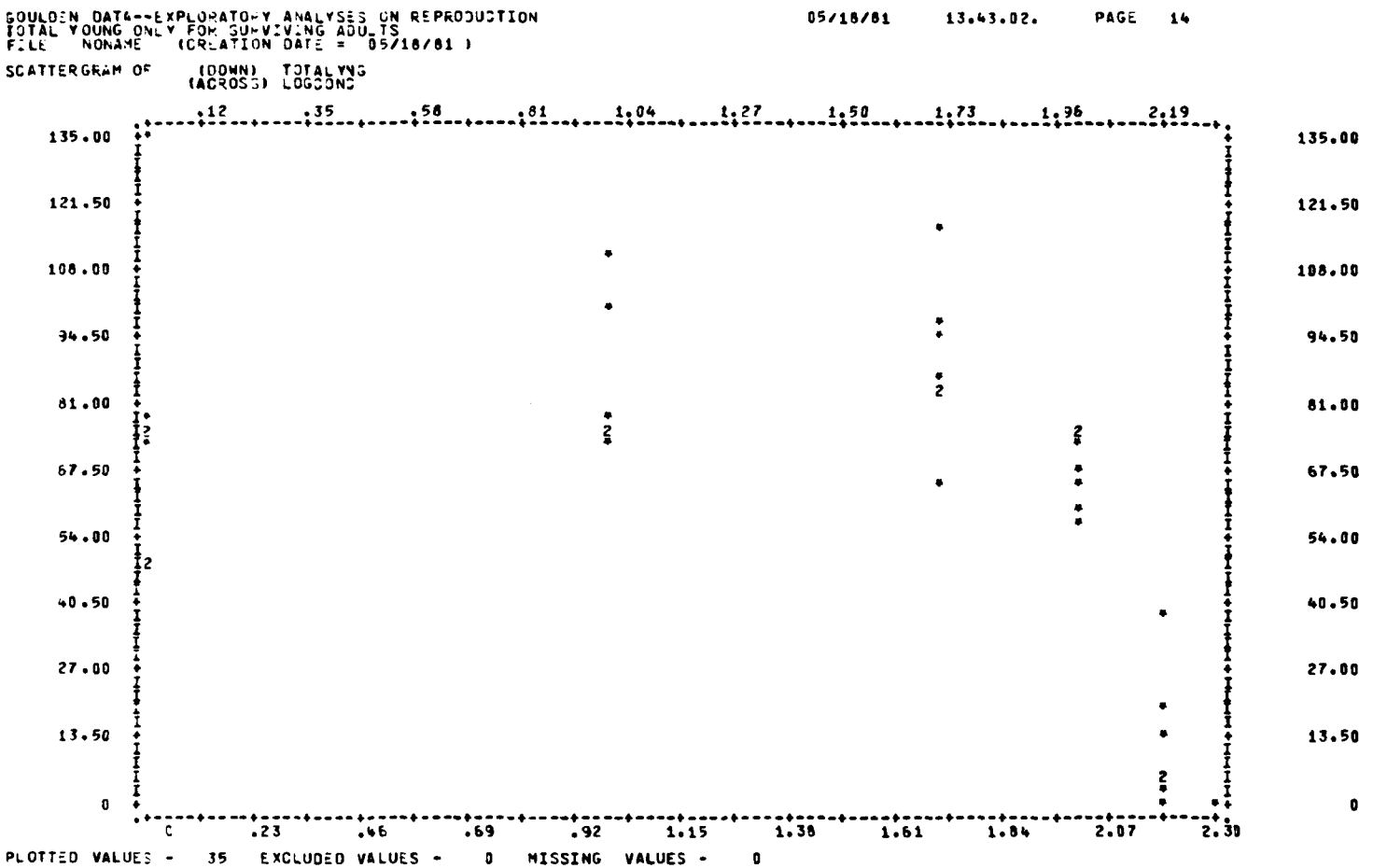


Figure II.23. Goulden - Isophorone. Individual 21 day production for surviving adults by $\log_{10}$ (concentration)

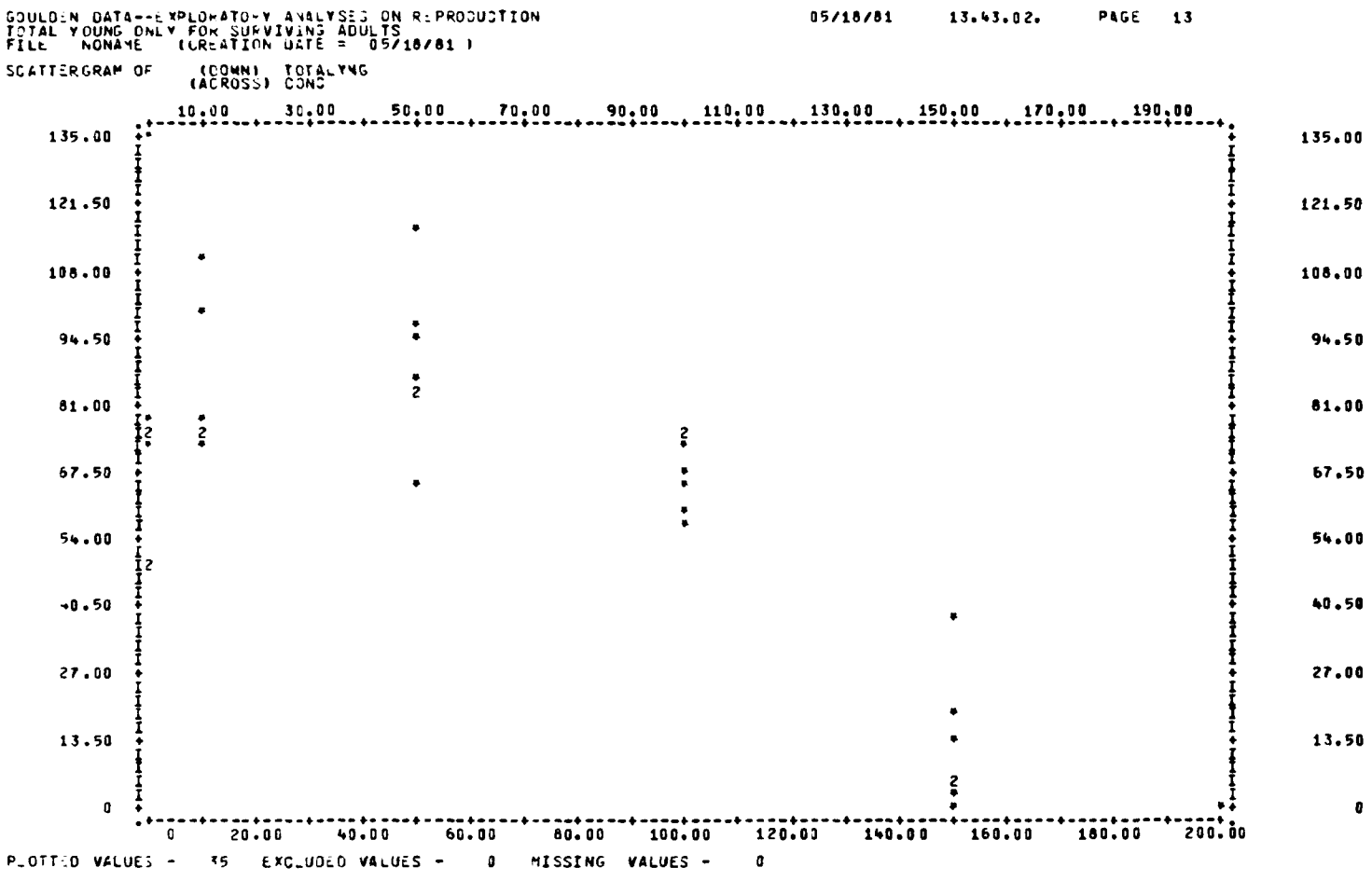
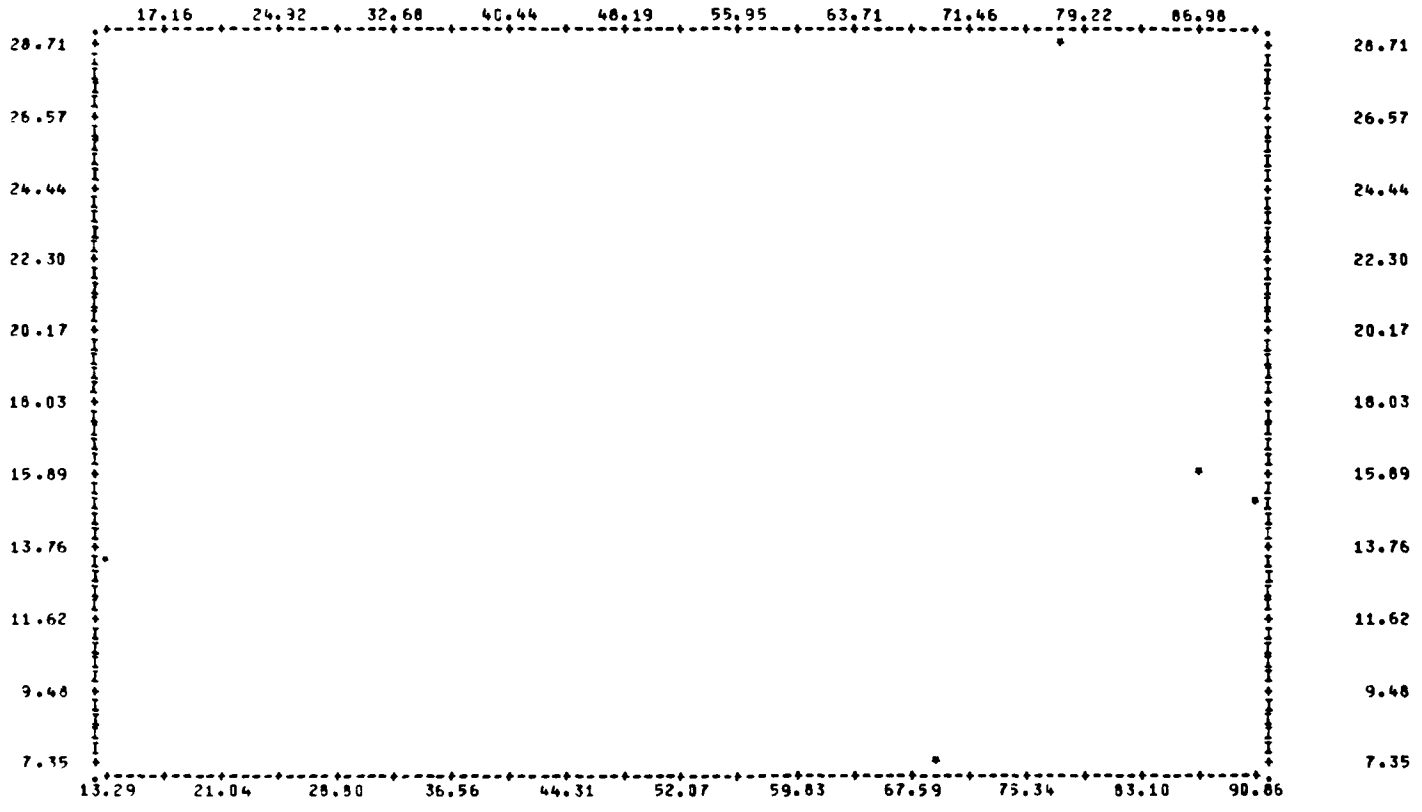


Figure II.24. Goulden - Isophorone. Individual 21 day production for surviving adults by concentration

GOULDEN REPRODUCTION--EXAMINING HETEROGENITY OF GROUP VARIANCES 05/18/81 13.43.12. PAGE 2

FILE NONAME (CREATION DATE = 05/18/81)

SCATTERGRAM OF (DOWN) STDDEV
(ACROSS) MEANYNG



STATISTICS..

CORRELATION (R)-	.23654	R SQUARED -	.05595	SIGNIFICANCE R -	.35083
STD ERR OF EST -	8.72789	INTERCEPT (A) -	12.31087	STD ERROR OF A -	10.14073
SIGNIFICANCE A -	.19550	SLOPE (B) -	.05851	STD ERROR OF B -	.13876
SIGNIFICANCE B -	.35083				

PLOTTED VALUES - 5 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure II.25. Goulden - Isophorone. Standard deviation of 21 day production for surviving adults by average production

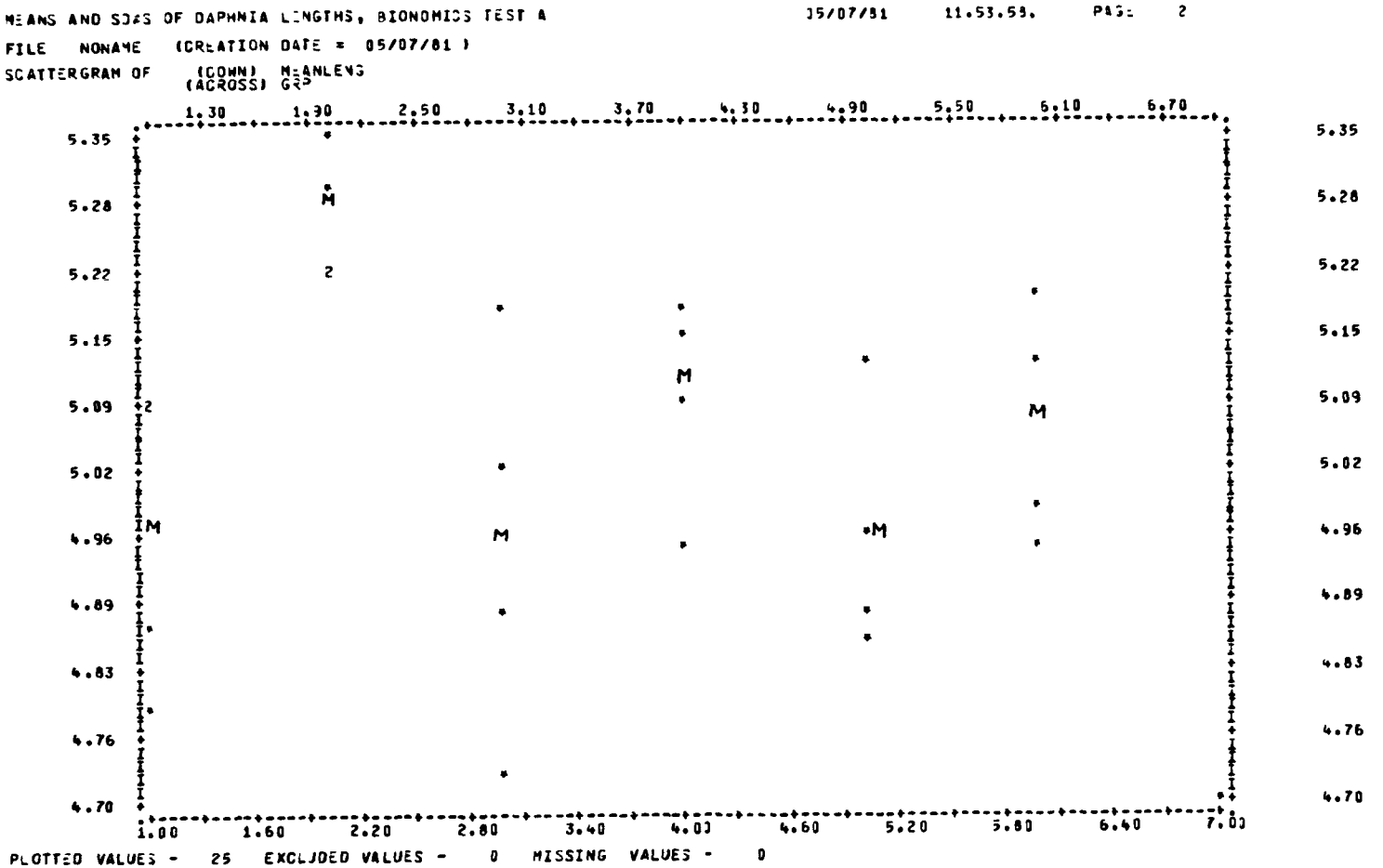
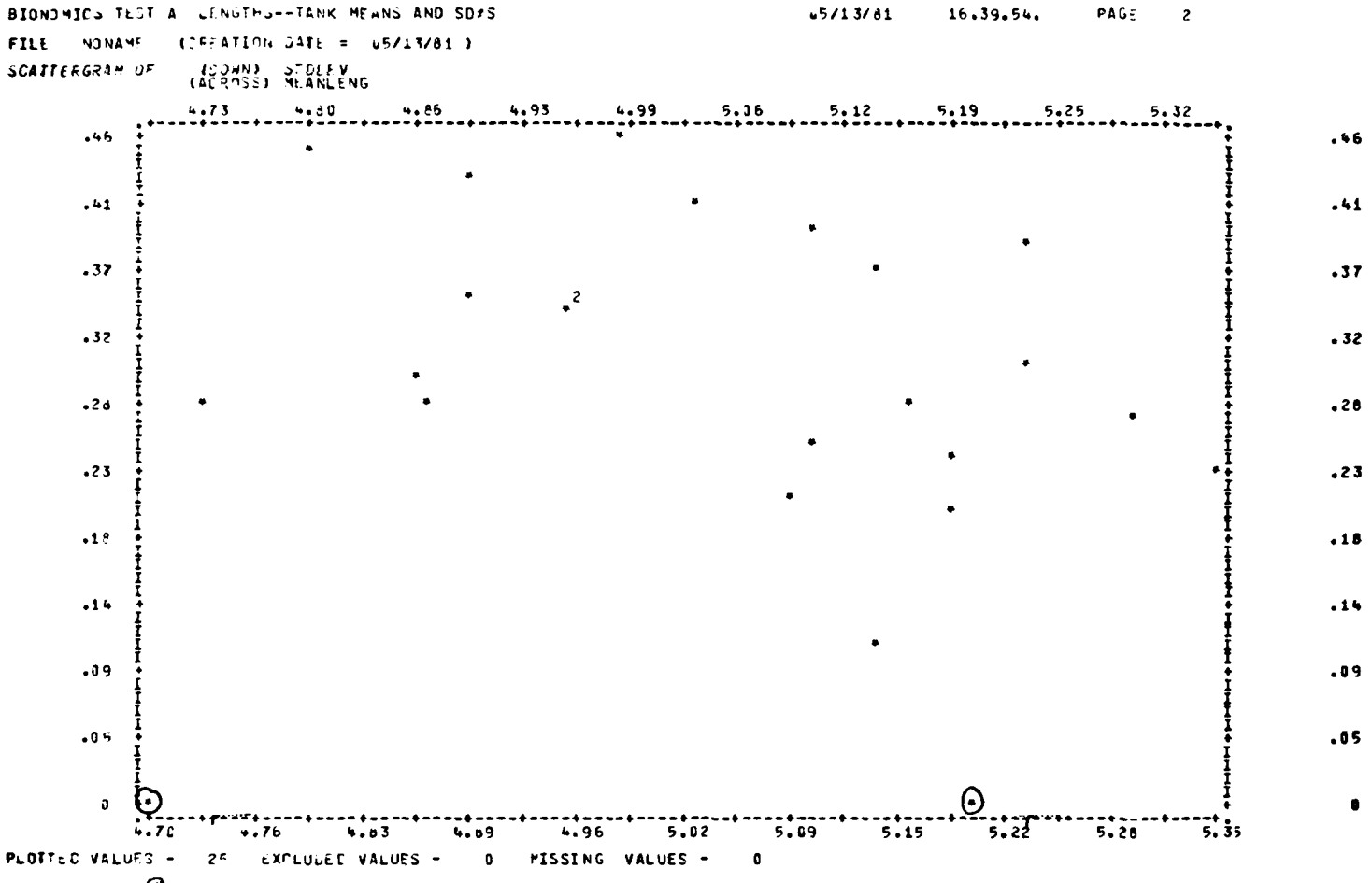


Figure II.26. LeBlanc - Test A. Average 21 day lengths of survivors within beakers by group. M's correspond to means of beaker averages.



Ⓐ based on just one surviving daphnid

Figure II.27. LeBlanc - Test A. Standard deviations of 21 day lengths of survivors within beakers by mean lengths within beakers

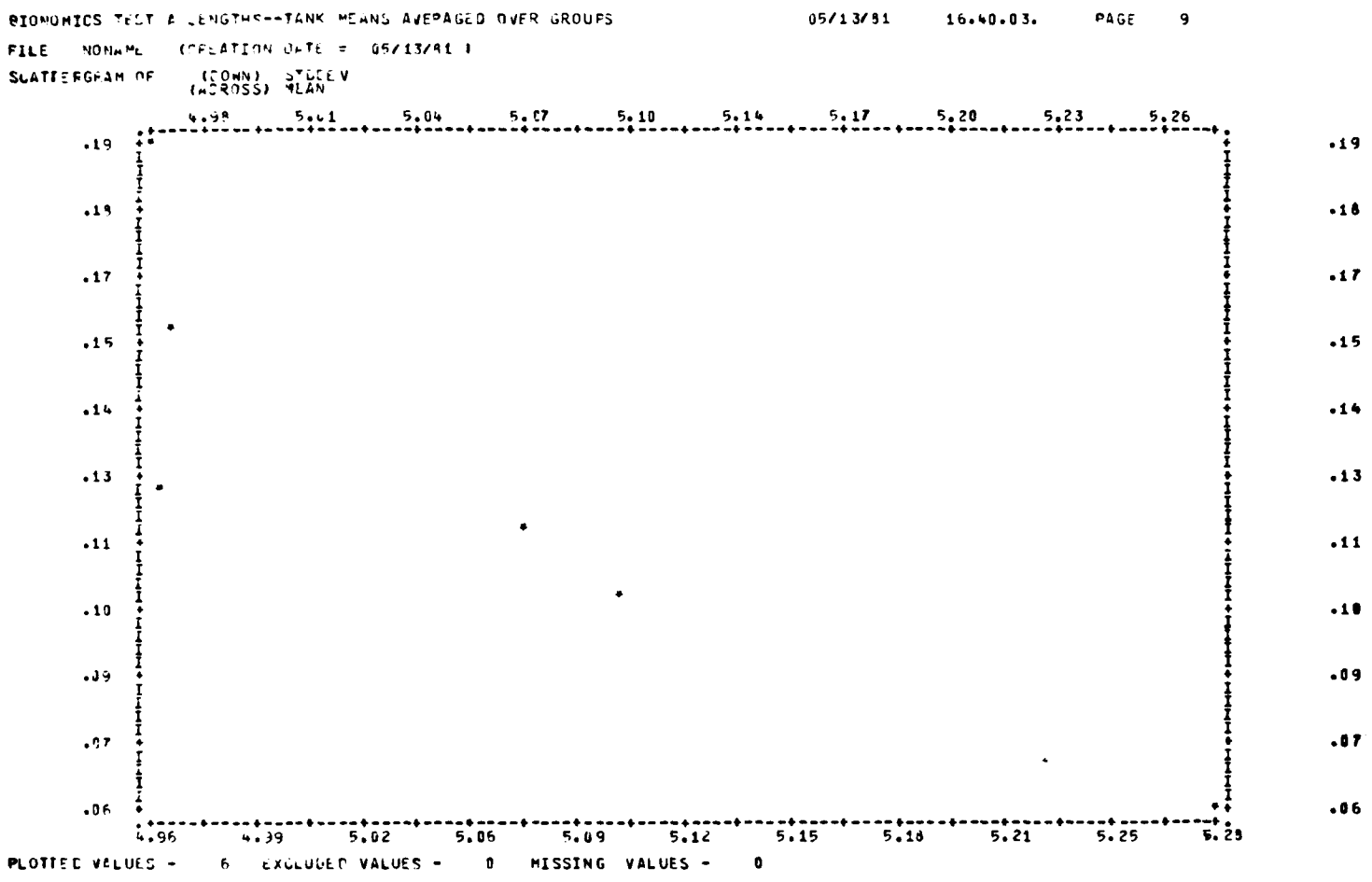


Figure II.28. LeBlanc - Test A. Standard deviations of average 21 day lengths among beakers within groups by means of beaker averages within groups (note that group 7 is omitted since just one beaker had a survivor)

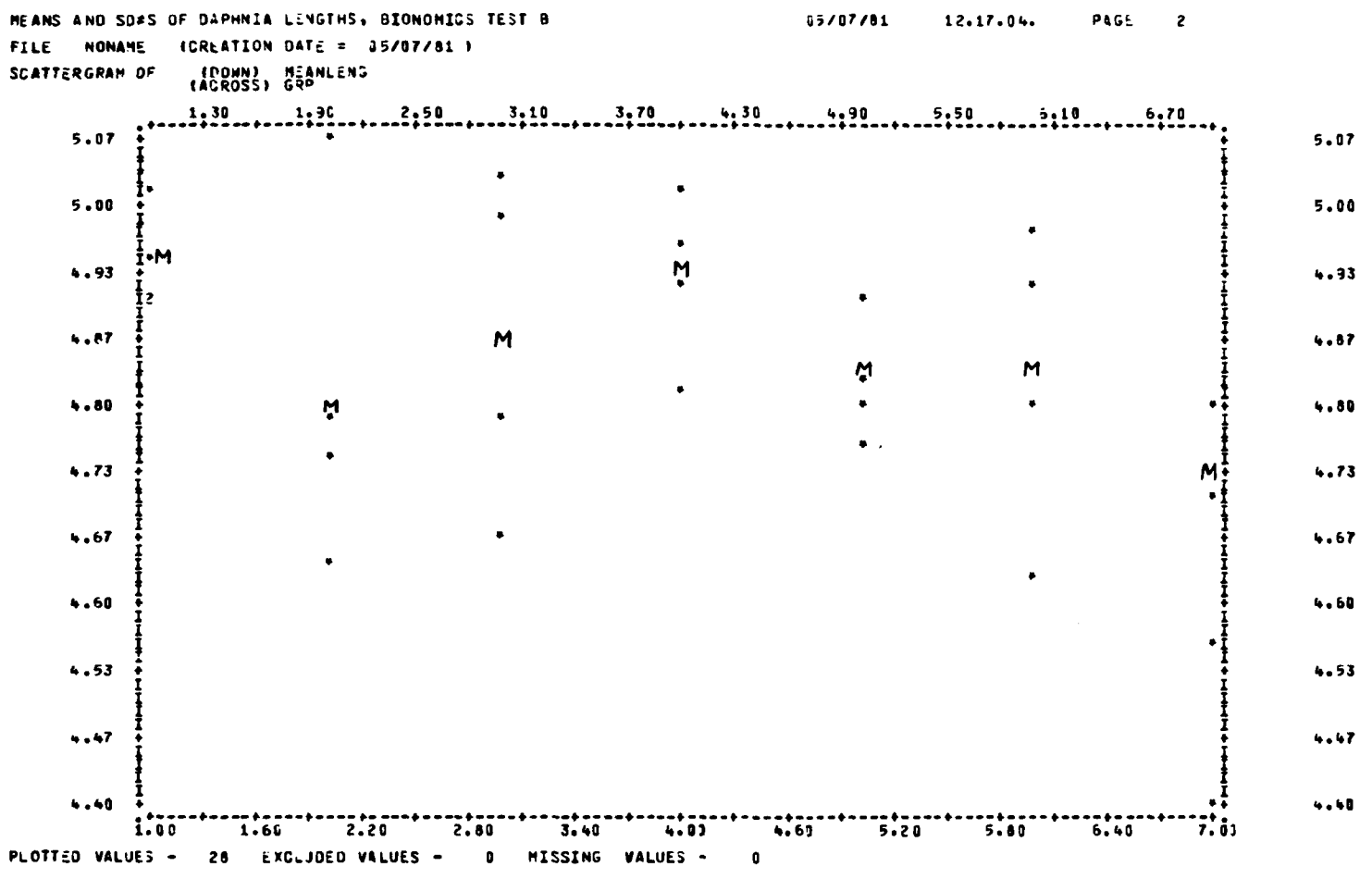


Figure II.29. LeBlanc - Test B. Average 21 day lengths of survivors within beakers by group. M's correspond to means of beaker averages

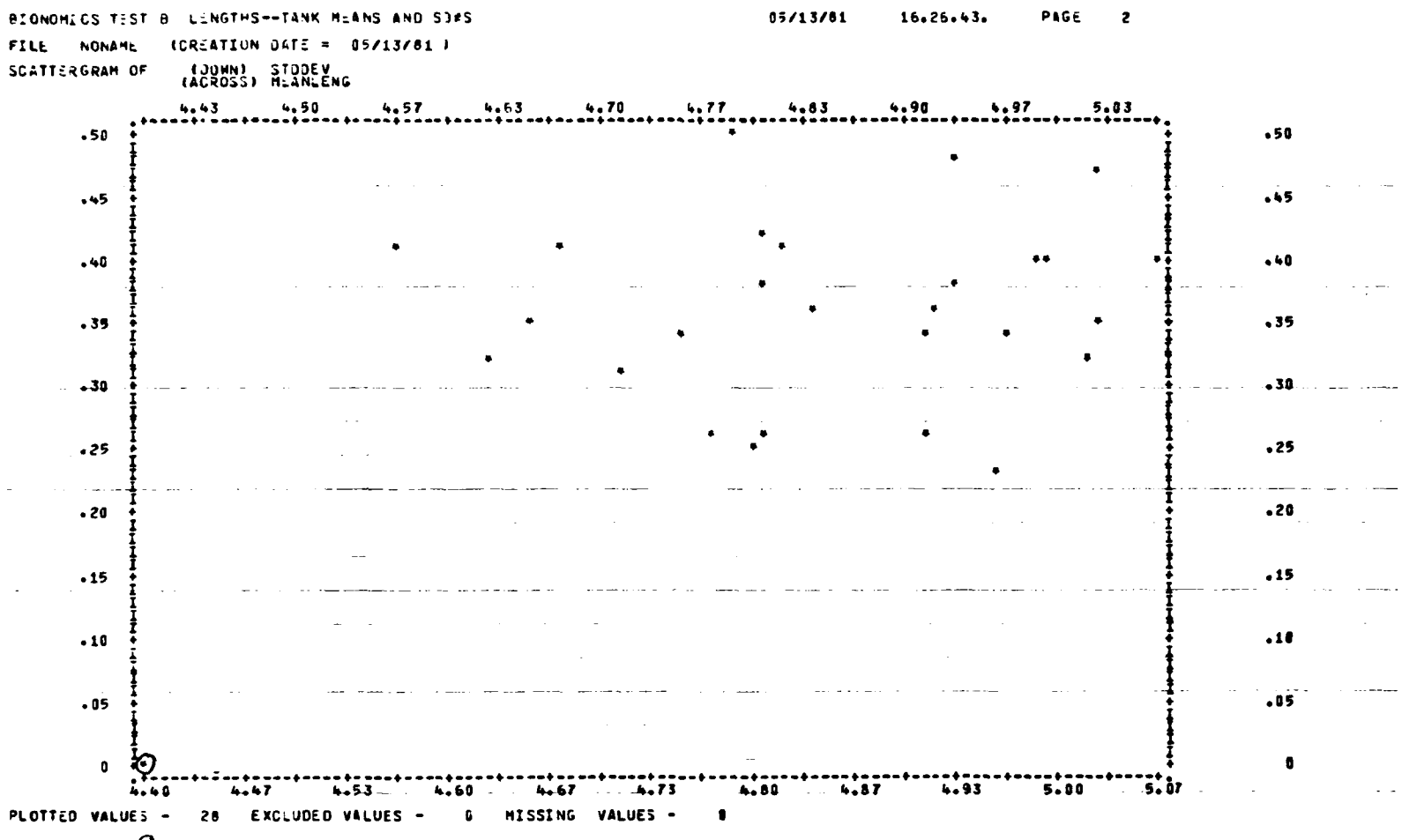


Figure II.30. LeBlanc - Test B. Standard deviations of 21 day lengths of survivors within beakers by mean lengths within beakers

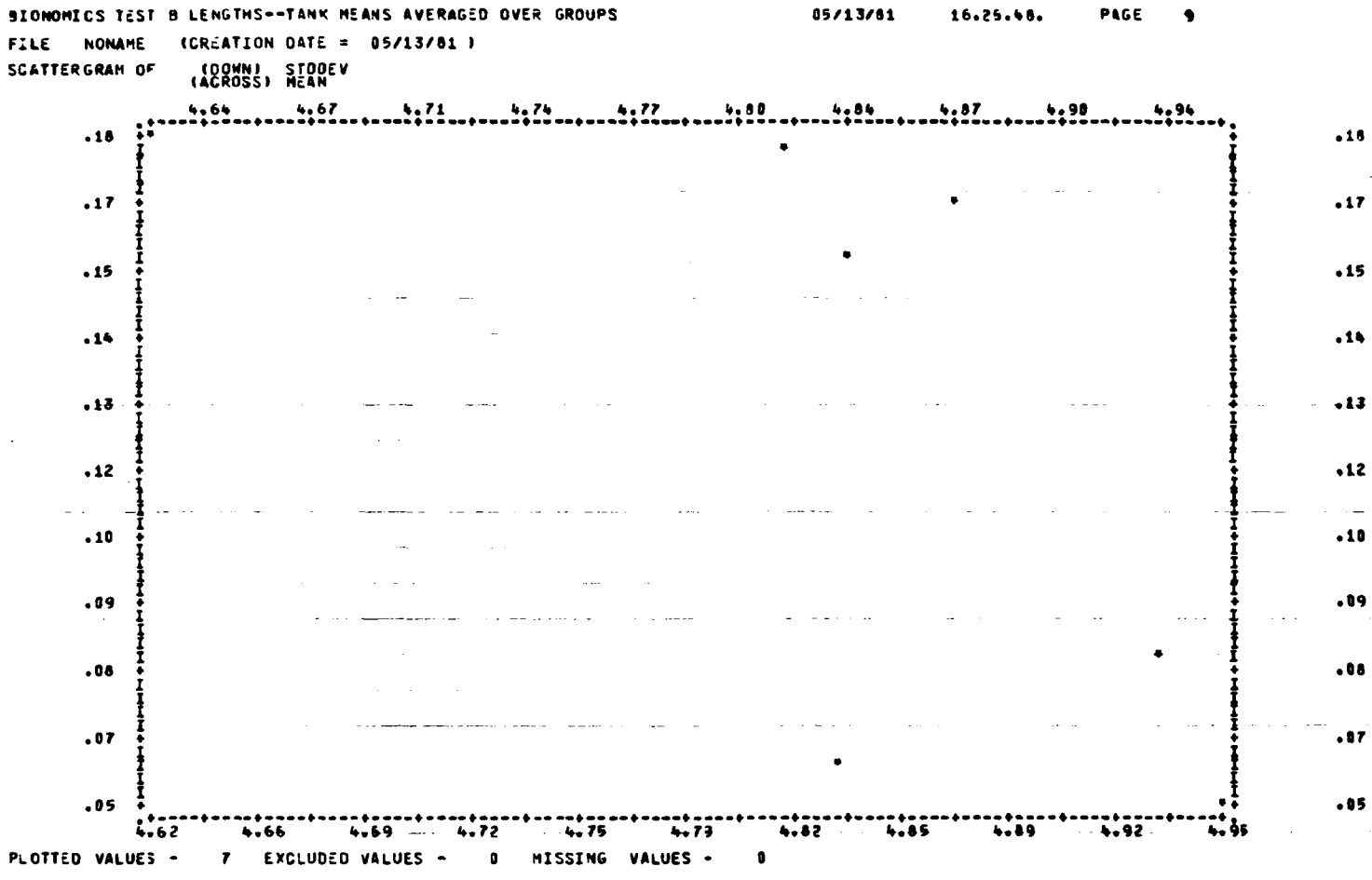


Figure II.31. LeBlanc - Test B. Standard deviations of 21 day lengths among beakers within groups by means of beaker averages within groups

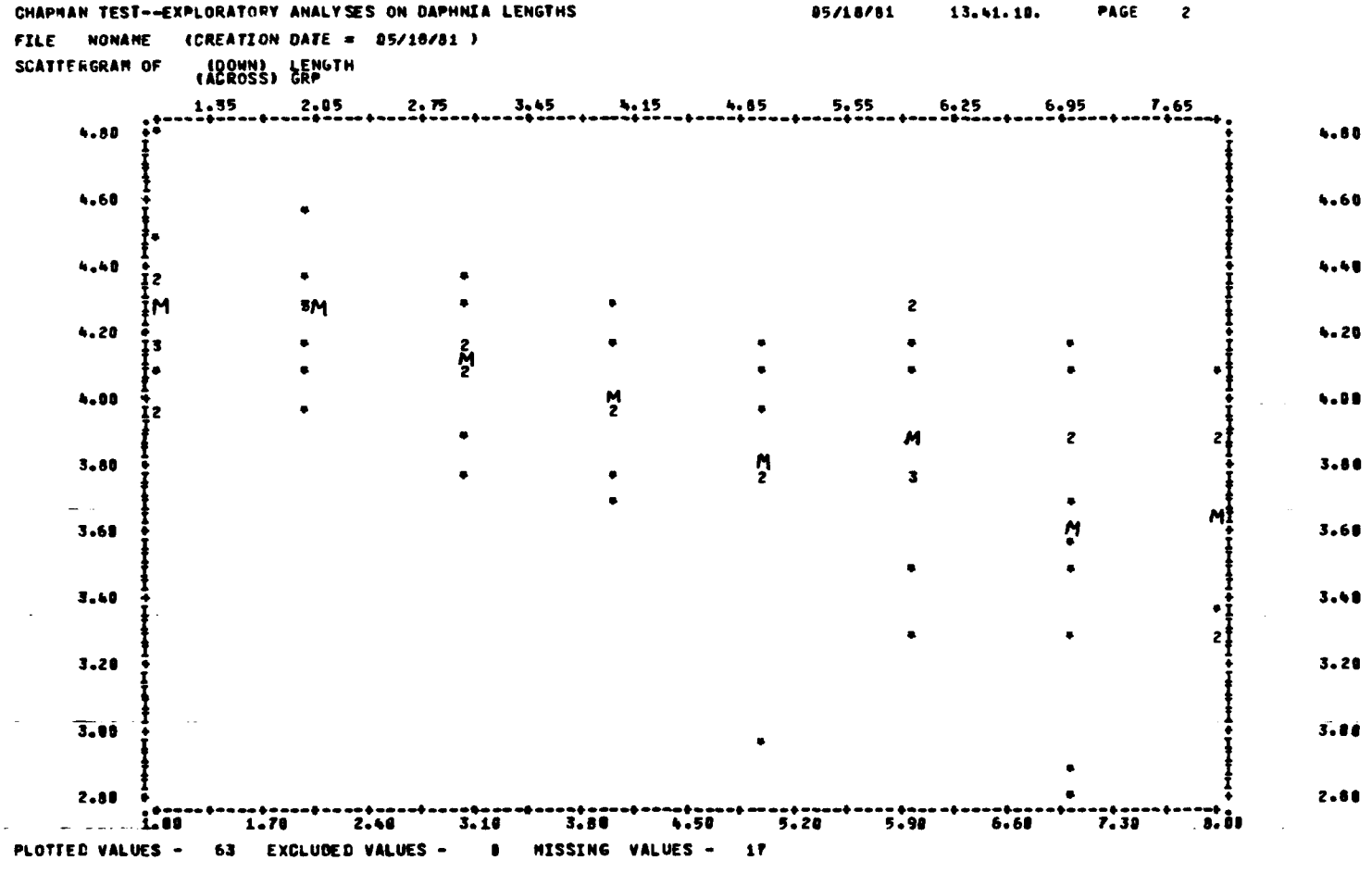


Figure II.32. Chapman - Beryllium. 21 day lengths of survivors within beakers by group. M's correspond to group means.

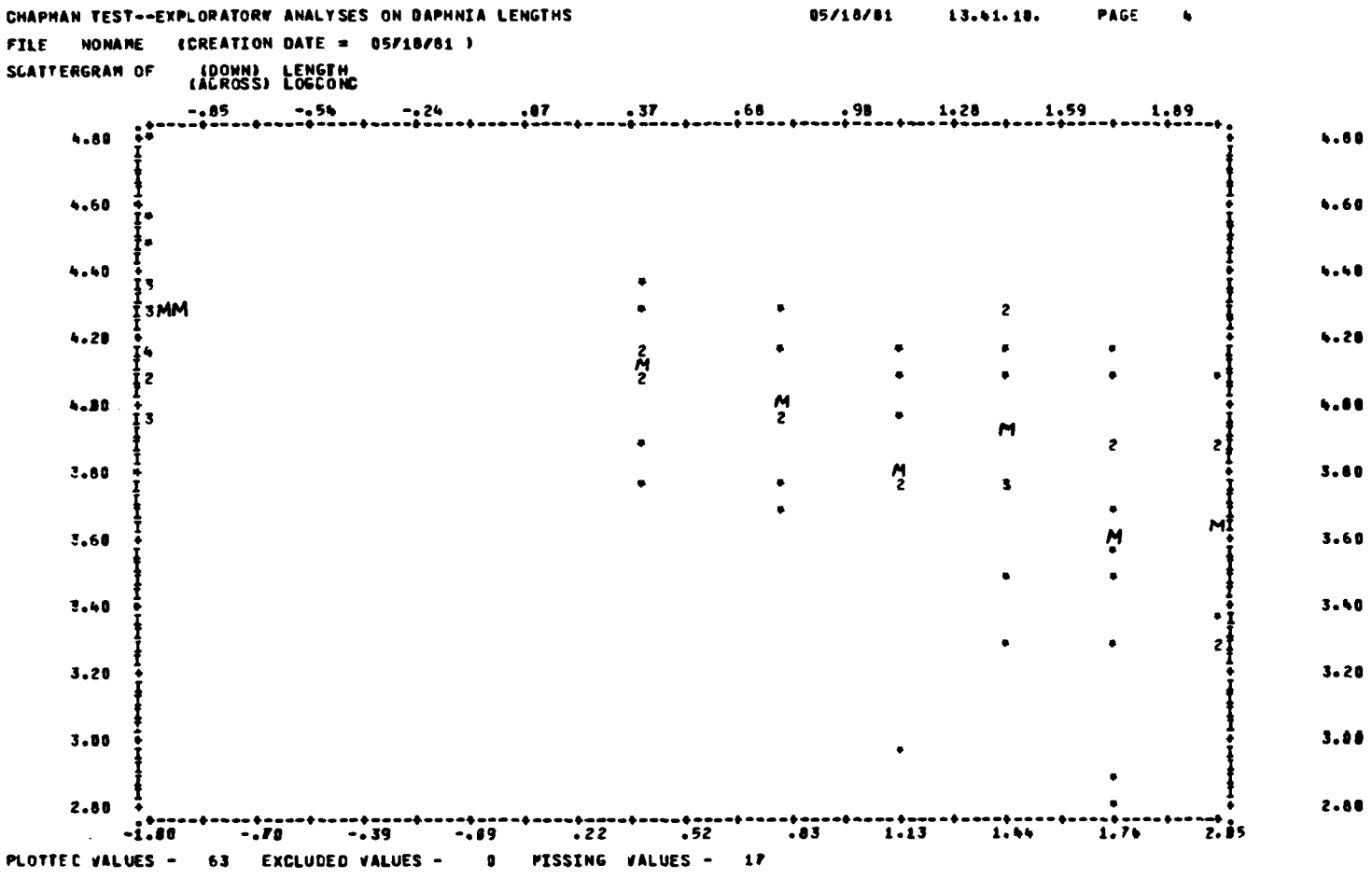
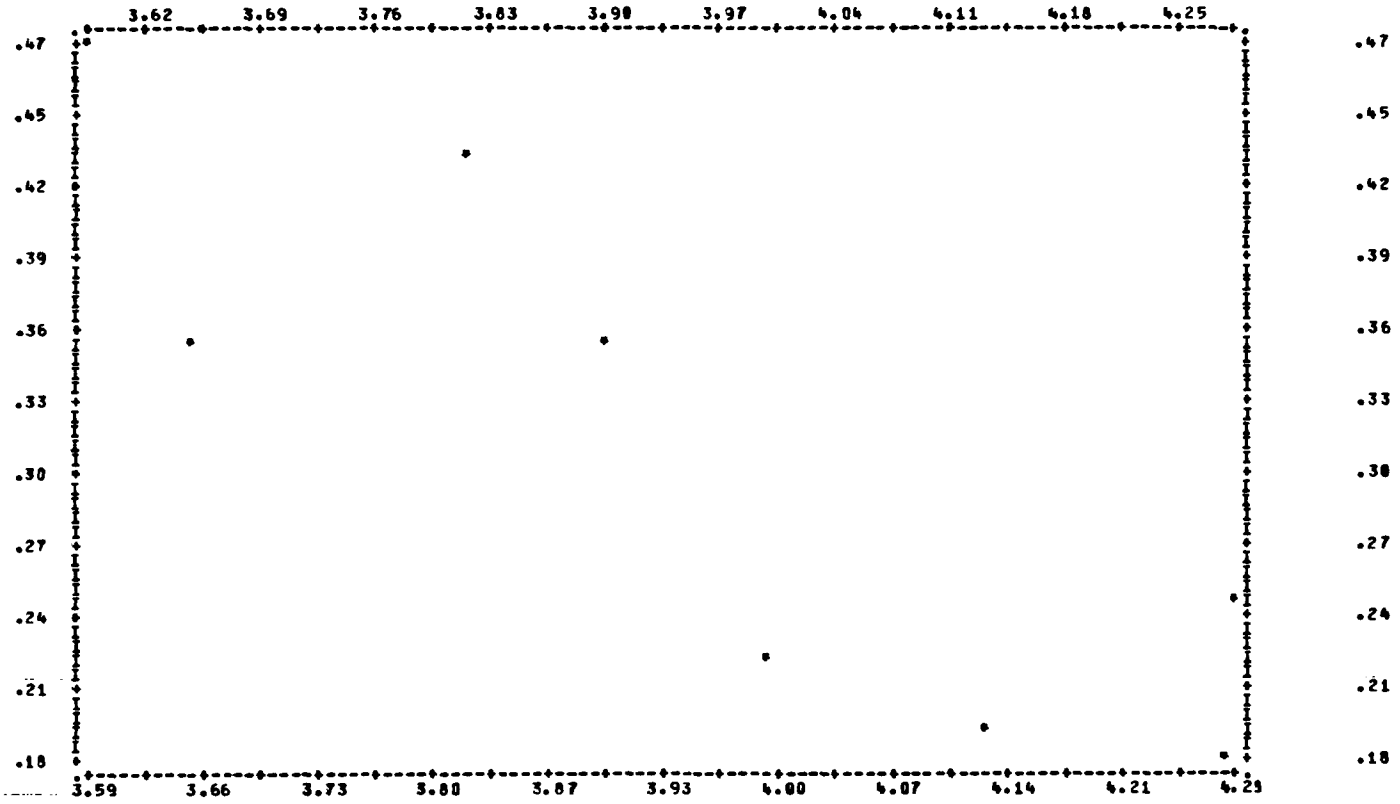


Figure II.33. Chapman - Beryllium. 21 day lengths of survivors within beakers by $\log_{10}$ (concentration). M's correspond to group means.

CHAPMAN LENGTHS--EXAMINING HETEROGENEITY OF GROUP VARIANCES 05/18/81 13.41.38. PAGE 2

FILE NONAME (CREATION DATE = 05/18/81)

SCATTERGRAM OF (DOWN) STDDEV
(ACROSS) MEANLENG



STATISTICS..

CORRELATION (R) -	-.86853	R SQUARED -	.75434	SIGNIFICANCE R -	.00257
STD ERR OF EST -	.05895	INTERCEPT (A) -	1.74261	STD ERROR OF A -	.33460
SIGNIFICANCE A -	.00100	SLOPE (B) -	-.36247	STD ERROR OF B -	.08445
SIGNIFICANCE B -	.00257				

PLOTTED VALUES - 8 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure II.34. Chapman - Beryllium. Standard deviations of 21 day lengths of survivors within groups by average lengths within groups

III. TESTING FOR BEAKER TO BEAKER HETEROGENEITY WITHIN TREATMENT GROUPS - SURVIVAL DATA

A. BACKGROUND

Toxicity tests with Daphnia magna generally include several replicate beakers per treatment or control group in order to be able to assess variability of response. An important preliminary inference is to determine if there is any statistical evidence of variation in response rate across beakers within groups. For survival data, it is of interest to determine whether the variation in mortality rates across beakers within groups is compatible with that to be expected based on binomial theory or whether it is in excess of that. In the former case the data may be pooled across beakers and subsequent analyses can be carried out based on binomial theory (i.e. on a per daphnid basis). However if beaker to beaker variation exists then standard errors based on binomial theory will underestimate the true variability of responses. This would lead to hypothesis tests that falsely reject the null hypothesis more often than the nominal level (i.e. inflated type 1 error), confidence intervals that are too short (i.e. attained confidence level lower than nominal), and simultaneous inference procedures arriving at no effect levels that are too low.

To account for possible beaker to beaker heterogeneity, variability estimates are generally based on sample variances of the observed survival rates in the replicate beakers with groups. Such variability estimates are appropriate whether or not beaker to beaker variation exists, however they are based on relatively few degrees of freedom and so reduce the sensitivity of subsequent analyses in the event that there is in fact no extra binomial variation. There is thus a tradeoff between possible underestimation of variability on the one hand and possible loss of sensitivity on the other. A reasonable compromise procedure is to first carry out a test for beaker to beaker heterogeneity among beakers within groups. If the test accepts the null hypothesis of no heterogeneity, we pool data across beakers and base subsequent analyses on binomial theory (i.e. per daphnid analysis). If the test rejects the null hypothesis of no heterogeneity, we either base subsequent analyses on sample response rates within each beaker (i.e. per beaker analyses) as is usually done or else adjust the sample sizes downward to effective sample sizes and then pool the adjusted data across beakers. The latter approach has been described and illustrated for fathead minnow data in Feder and Collins [1], Sections VII, VIII and IX and will be further illustrated in this report for Daphnia magna data. In this section we confine attention to testing for heterogeneity among beakers. We discuss adjustment procedures in subsequent sections.

Feder and Collins [1] Sections VII and VIII remark that problems can arise with some of the "standard" procedures for testing for beaker to beaker variation within groups. These tests are based on asymptotic theory and sometimes are based on a specific form of dose response model. There are two possible types of difficulties with the usual asymptotic chi square

tests of goodness of fit. First, the weights used in the denominator of the chi square statistic are inappropriate if the assumed form of dose response model is inappropriate. This may bias results. For example a chi square test statistic for heterogeneity among beakers within groups based on a model of constant mortality rates across treatment groups has constant probability weights across groups in the denominator and thus would be inappropriate if in fact there is a trend in mortality rate with increasing group number. Similarly a chi square test statistic for heterogeneity based on a probit model has probit based weights in the denominator and so would be inappropriate if the probit model is inappropriate, etc.

A second possible problem with the "standard" heterogeneity tests results from their asymptotic nature. The validity of the asymptotic chi square theory on which they are based is dependent on expected response frequencies being large enough. If just a single response is observed in a group with very small expected frequency, the contribution of that group to the overall chi square value can be dominant and can strongly bias the resulting test of heterogeneity among beakers. This situation was demonstrated by Feder and Collins, Tables VII.1 and VII.2. Thus the relatively small sample sizes coupled with response rates close to 0 or 1, that are fairly common in aquatic toxicity tests, result in small expected frequencies and therefore often invalidate the assumptions underlying heterogeneity tests based on asymptotic theory

To account for these two problems, Feder and Collins carry out separate chi square heterogeneity tests within each concentration group without imposing any structure on the form of the concentration-response relation. The test results are then pooled across groups to result in an overall test. The tests are based either on asymptotic theory or on exact, small sample theory depending on whether the expected response frequencies within each cell are large or small. The convention we have used has been to use heterogeneity tests based on exact, small sample theory if any expected response frequencies are less than 5. A computer program, EXAX2, has been developed to carry out tests of heterogeneity of mortality rates among beakers within groups, based on exact, small sample theory. This program is described in detail and illustrated with data based on fathead minnow tests in Feder and Collins, Section VIII and in Feder and Willavize [2]. See those reports for details of the program. In this section we illustrate the use of EXAX2 on several sets of data from toxicity tests with Daphnia magna to test for beaker to beaker heterogeneity in mortality response rates within groups.

B. APPLICATION OF EXAX2 TO TEST FOR HETEROGENEITY AMONG BEAKERS WITHIN GROUPS WITH DAPHNIA MAGNA DATA

In this subsection we illustrate the results of EXAX2 comparisons of 21 day mortality rates among replicate beakers within groups using data from a number of toxicity tests. The EXAX2 outputs are shown in the referenced figures. The observed and expected cell frequencies are indicated. If any of the expected cell frequencies are lower than the (user specified) cutoff of 5, exact distribution theory is used for the comparison in that group. The exact distribution of the chi square statistic,

conditional on the marginal totals, is enumerated and displayed. The observed value of chi square, the observed significance level A_i $-2\ell n\ A_i$, $E(-2\ell n\ A_i)$, $Var(-2\ell n\ A_i)$ are calculated. The separate independent tests for each treatment or control group are combined by summing $-2\ell n\ A_i$, $E(-2\ell n\ A_i)$, and $Var(-2\ell n\ A_i)$ over groups and calculating the standardized test statistic, Z, which is then compared to a standard normal distribution.

We now illustrate this procedure.

LeBlanc Test A 21 day Mortality

There are a water control group, a solvent control group, and five treatment groups. There are four replicate test chambers per treat group, 20 daphnids per chamber to start. The results from the EXAX2 calculations are shown in Figures III.1 to III.7 and are summarized below. The mortality results are displayed graphically in Figure II.1. It is obvious from Figure II.1 that there is great heterogeneity of mortality rates in group 6 and that group 3 possibly exhibits some heterogeneity. Other than that, the results appear to be homogeneous across beakers within groups. These conclusions from Figure II.1 are supported by the results in the EXAX2 output. Figures III.1, III.2, III.4, III.5 and III.7 show no significant heterogeneity of mortality across beakers within groups 1, 2, 4, 5 and 7 respectively. Figure III.3 shows marginal evidence of heterogeneity across beakers, with beakers 3 and 4 having somewhat greater mortality than beakers 1 and 2. Figure III.6 shows substantial heterogeneity of response across beakers. Beakers 3 and 4 have better than 75 percent mortality while beakers 1 and 2 have less than 20 percent mortality. This is of course both highly statistically and highly biologically significant. The pooled significance level calculations are presented below the results for group 7. Z is highly statistically significant since the probability of a standard normal deviate exceeding 5.553 by chance is essentially 0.

LE BLANC TEST A 21 DAY MORTALITY

Trt	Method	Chi Sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	0.0000	1.0000	0.0000	1.6609	3.6367
2	exact	3.8095	0.6105	0.9869	0.8687	1.6419
3	exact	7.7714	0.0757	5.1619	1.5750	3.4893
4	exact	4.2270	0.3445	2.1312	1.5068	3.1226
5	exact	5.4795	0.1975	3.2436	1.5068	3.1226
6	asymptotic	41.2000	0.0000	37.8863	2.0000	4.0000
7	exact	3.0380	1.0000	0.0000	0.0000	0.0000

$\Sigma\ -2\ell n\ A_i = 49.410$ $\Sigma\ E(-2\ell n\ A_i) = 9.1183$ $\Sigma\ Var\ (-2\ell n\ A_i) = 19.0129$
$Z = 5.553$

The highly significant pooled result is due to group 6. Without this group, the value of Z would be just 1.00. We thus conclude that there is overall strong statistical evidence of beaker to beaker heterogeneity within groups and this is due primarily to the strong dichotomy of mortality results in group 6.

It is interesting to note that in both group 3 and group 6, where dichotomies in mortality exist, beakers numbered 3 and 4 exhibit somewhat greater mortality rates than beakers numbered 1 and 2. One cannot help but wonder whether this is a coincidence or whether there is some systematic difference between beakers having different numbers. This could be related to source of test daphnids, placement or handling of beakers, connections to proportional diluter, or some other factors associated with experimental technique. These issues should be discussed in detail with the investigator and any systematic effects should be taken into account in subsequent analyses and interpretations. However we leave this issue here and do not pursue it in this study.

LeBlanc Test B 21 day Mortality

The set up of the test is the same as that for Test A, namely water and solvent control groups, five treatment groups, four beakers per group, 20 daphnids per beaker to start. The results from the EXAX2 calculations are summarized below. The mortality results are displayed graphically in Figure II.3. The pattern of results is remarkably like that in Test A. It is obvious from Figure II.3 that group 7 exhibits the same kind of substantial dichotomy of mortality rates that group 6 exhibited in Test A. Some of the other groups show some suggestions of heterogeneity.

LE BLANC TEST B 21 DAY MORTALITY

Trt	Method	Chi sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	7.7778	0.0711	5.2879	1.5361	3.4424
2	exact	1.0103	0.9089	0.1911	1.7273	3.6654
3	exact	4.2105	0.3222	2.2652	1.1245	2.3815
4	exact	5.7600	0.1591	3.6768	1.3590	2.6405
5	exact	1.7451	0.7153	0.6702	1.7273	3.6654
6	exact	4.4444	0.2521	2.7562	1.5361	3.4424
7	asymptotic	26.6667	0.0000	23.7637	2.0000	4.0000

$\Sigma\ -2\ell n\ A_i = 38.6110$ $\Sigma\ E(-2\ell n\ A_i) = 11.0104$ $\Sigma\ Var(-2\ell n\ A_i) = 23.2375$
$Z = 3.9863$

The pooled significance level calculations are presented below those for group 7. Z is highly statistically significant since the probability of a standard normal deviate exceeding 3.9863 by chance is about 0.0001. The highly significant pooled result is due to group 7. Without this group the value of Z would be just 1.165. We thus conclude that there is overall strong statistical evidence of beaker to beaker heterogeneity within groups and this is due primarily to the strong dichotomy of mortality results in group 7.

It is curious that the patterns of beaker to beaker heterogeneity are so similar in Tests A and B. Each test has one group with a very substantial dichotomy of mortality rates, while the other groups demonstrate little or no heterogeneity.

Adams-Selenium 21 Day Mortality

This test is a static renewal test. There are a control group and seven treatment groups. There are three replicate beakers per group and five daphnids per group to start. The results from the EXAX2 calculations are summarized below. The mortality results are displayed in Figure II.5. We see that in each group there is nearly no mortality or complete mortality. The middle portion of the dose response curve undoubtedly lies between the concentrations in groups 4 and 5. There is of course no suggestion of heterogeneity within groups.

ADAMS-SELENIUM 21 DAY MORTALITY

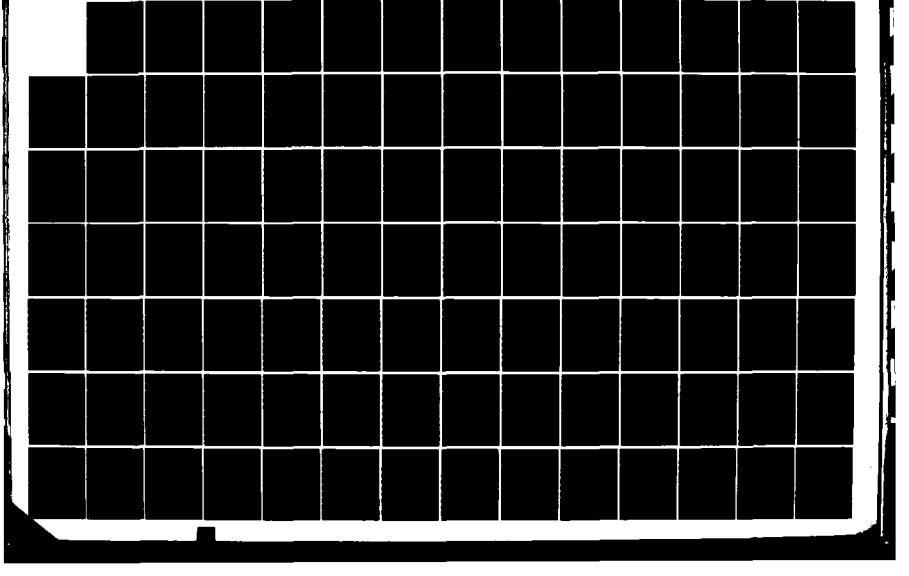
Trt	Method	Chi Sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	2.14286	1.000	0.0000	0.0000	0.0000
2	table degenerate		1.000	0.0000	0.0000	0.0000
3	exact	2.14286	1.000	0.0000	0.0000	0.0000
4	table degenerate		1.000	0.0000	0.0000	0.0000
5	table degenerate		1.000	0.0000	0.0000	0.0000
6	table degenerate		1.000	0.0000	0.0000	0.0000
7	table degenerate		1.000	0.0000	0.0000	0.0000
8	table degenerate		1.000	0.0000	0.0000	0.0000

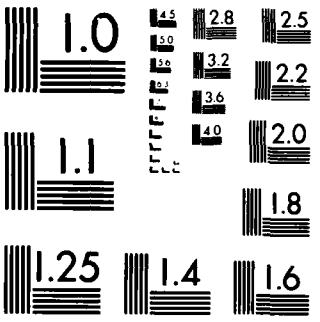
$\Sigma\ -2\ell n\ A_i = 0.000$ $\Sigma\ E(-2\ell n\ A_i) = 0.000$ $\Sigma\ Var(-2\ell n\ A_i) = 0.000$
Z is indeterminate

In brief we see no beaker to beaker heterogeneity within groups since there is essentially either no mortality or complete mortality within each group.

AD-A152 070 DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA(U) BATTELLE COLUMBUS LABS OH P I FEDER DEC 81 BATT-G-7673 DAMD17-80-C-0165 2/5

UNCLASSIFIED F/G 6/20 NL





Goulden-Isophorone 21 Day Mortality

This test is also a static renewal test. There are a control group and five treatment groups. Each group consists of 10 beakers. Beakers 1-7 contain just individual daphnids. These are intended primarily to measure productivity. Beakers 8-10 contain five daphnids each. These are intended to estimate survival. Several different EXAX2 runs were carried out In one run, mortality rates in beakers 8, 9 and 10 were compared against one another separately within each group. This is to determine whether there is any evidence of beaker to beaker heterogeneity among the multiple daphnid beakers within groups. In another EXAX2 run mortality was pooled across beakers 1-7 and this pooled mortality was compared with mortality pooled across beakers 8, 9 and 10 within each group. This is to determine whether there is any evidence of differences in mortality rates between singly housed and multiply housed daphnids within each dose group. The survival results are displayed graphically in Figure II.8. The pooled survival rates for the individually housed daphnids are plotted with the plotting symbol "7". The survival rates for the beakers with multiple daphnids are plotted with the plotting symbol "*". If two multiple daphnid survival rates coincide, their common value is plotted with a "2". If one or two "*"'s coincide with a "7", their common value is plotted with an "8" or "9". We see that with the exception of group 6, there is just one death among singly housed daphnids. The multiply housed daphnids appear to have somewhat greater mortality. The results from the EXAX2 calculations comparing mortality rates among the multiply housed daphnids are summarized below.

GOULDEN-ISOPHORONE 21 DAY MORTALITY - COMPARISONS
AMONG BEAKERS WITH MULTIPLE DAPHNIDS

Trt	Method	Chi Sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	1.1539	1.0000	0.0000	0.7159	1.2812
2	exact	2.1429	1.0000	0.0000	0.0000	0.0000
3	exact	2.1429	1.0000	0.0000	0.0000	0.0000
4	exact	2.5000	0.7253	0.6424	0.7821	1.6103
5	exact	6.9643	0.0676	5.3883	1.2813	2.9369
6	exact	1.1539	1.0000	0.0000	0.7159	1.2812

$\Sigma\ -2\ell n\ A_i = 6.0307$ $\Sigma\ E(-2\ell n\ A_i) = 3.4952$ $\Sigma\ Var(-2\ell n\ A_i) = 7.1095$
$Z = 0.8221$

The probability that a standard normal random variable exceeds 0.82 is about 0.20. We see that there is thus no statistical evidence of heterogeneity among beakers within groups. Most of the heterogeneity that is observed is that from group 5, where there is marginal suggestion of beaker to beaker heterogeneity.

We now pool mortality results across beakers with multiple daphnids and compare these with mortality results for the individually housed daphnids. There are thus two subgroups per treatment group. The mortality rates for the individually housed daphnids are based on a sample of size 7 while the mortality rates for the multiply housed daphnids are based on a sample of size 15. The EXAX2 output is given in Figures III.8 to III.13 and the results are summarized below.

GOULDEN-ISOPHORONE 21 DAY MORTALITY - COMPARISONS
BETWEEN POOLED MORTALITY RATES FOR INDIVIDUALLY HOUSED DAPHNIDS
AND POOLED MORTALITY RATES FOR MULTIPLY HOUSED DAPHNIDS

Trt	Method	Chi Sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	1.0267	0.5455	1.2123	0.9870	1.7847
2	exact	0.3352	1.0000	0.0000	0.9870	1.7847
3	exact	0.4889	1.0000	0.0000	0.7287	1.1379
4	exact	1.6211	0.5227	1.2974	1.1614	2.2462
5	exact	5.8667	0.0225	7.5912	1.4030	3.0042
6	exact	0.0037	1.0000	0.0000	1.1614	2.2462

$\Sigma\ -2\ell n\ A_i = 10.1009$ $\Sigma\ E(-2\ell n\ A_i) = 6.4286$ $\Sigma\ Var(-2\ell n\ A_i) = 12.2040$
$Z = 0.9329$

The probability of a standard normal random variable exceeding 0.9329 is 0.175. Thus except for group 5, there is no statistical evidence of differences in mortality rates between individually and multiply housed daphnids. The statistically significant result in group 5 is due to the single beaker in which all five daphnids died. Without this beaker the mortality results would be

	Live	Die	
Individual	7	0	7
Multiple	7	3	10
	14	3	17

An approximate two-tailed probability of observing as extreme a result just due to chance is, by the hypergeometric distribution,

$$2\,\frac{\binom{7}{0}\binom{10}{0}}{\binom{17}{3}} = 2\left(\frac{3}{17}\right) = 0.35 \quad .$$

We conclude that there is no statistical evidence of overall differences in mortality rates between individually and multiply housed daphnids. The interpretation of the significant difference in group 5 depends on the reason for complete mortality in one of the multiple beakers. It should be noted however that in four of the six groups the individually housed daphnids had lower mortality rates than the multiply housed daphnids and the rates were essentially the same in a fifth group. Thus a significant difference might have shown up had the sample sizes been greater.

The previous chi square based test is a two sided test. We can attain greater power for testing for effects if we carry out a one sided test that, for example, mortality is lower among individually housed daphnids than among multiply housed daphnids. One way to carry out a test of equality of mortality rates against a one sided alternative, based on asymptotic theory, is to perform an arc sin transformation and carry out a normal theory based test. Although the assumption of asymptotic normality is stretched a bit with the small sample sizes in this example, we carry out the test for illustrative purposes.

Within each group, the transform 2 arc sin $\hat{p}^{1/2}$ has an asymptotic normal distribution with mean 2 arc sin $p^{1/2}$ and standard deviation $1/n^{1/2}$. If singly and multiply housed daphnids have the same mortality rates then differences among their arc sin transforms will have mean 0.

	Singly Housed			Multiply Housed		
Group	$\hat{p}$	2 arc sin $\hat{p}^{1/2}$	Var	$\hat{p}$	2 arc sin $\hat{p}^{1/2}$	Var
1	0	0	0.143	0.133	0.748	0.067
2	0.143	0.775	0.143	0.067	0.522	0.067
3	0	0	0.143	0.067	0.522	0.067
4	0	0	0.143	0.20	0.927	0.067
5	0	0	0.143	0.533	1.638	0.067
6	0.857	2.366	0.143	0.867	2.394	0.067

Taking differences within each group yields

Group	DIFF (Mult - Single)	Std Err (DIFF)
1	0.748	0.458
2	-0.253	0.458
3	0.522	0.458
4	0.927	0.458
5	1.638	0.458
6	0.028	0.458

Under the null hypothesis, DIFF is asymptotically normal with mean 0 and std err 0.458. The average difference is 0.602 with a standard error of $0.458/6^{1/2} = 0.187$. Thus the one sided test demonstrates a statistically significantly greater mortality rate among the multiply housed daphnids then among the singly housed daphnids. If we exclude group 5 because of the beaker with no survivors, the average difference in the remaining five groups is 0.394 with a standard error of $0.458/5^{1/2} = 0.205$. The probability of a standard normal deviate exceeding $0.394/0.205 = 1.922$ is 0.027. Thus even without group 5, this one sided test suggests some statistical evidence for greater mortality among the multiply housed daphnids. However the validity of the asymptotic theory in this example is questionable. The validity of this one sided asymptotic test should be studied in greater detail.

In summary, the exact small sample heterogeneity test based on the chi square distribution reveals no overall heterogeneity among the mortality responses in the beakers with multiple daphnids. It also reveals no overall heterogeneity between average mortality per group for the individually and the multiply housed daphnids. The one sided, asymptotic test of heterogeneity reveals some statistical evidence of differences in mortality between the individually and multiply housed daphnics. The conclusions about differences in mortality between singly and multiply housed daphnids are thus tentative.

C. APPLICATION OF EXAX2 TO TEST FOR HETEROGENEITY AMONG BEAKERS WITHIN GROUPS AFTER ADJUSTING FOR EARLY LIFE STAGE MORTALITY

The mortality comparisons among beakers within groups discussed in the previous subsection were all based on 21 day mortality. This measure of mortality is an overall measure and encompasses both early and later stage mortality. In each of the data sets discussed in this study, mortality is measured at a number of points in time during the course of the test. LeBlanc measures mortality at 7, 14 and 21 days. Adams measures mortality at days 2, 4, 7, 9, 11, 14, 16, 18 and 21. Chapman and Goulden also measure mortality at nine time points during the test. These intermediate mortality responses can be used to separate inferences about overall mortality into inferences about early life stage mortality and later life stage mortality. One simple way of doing this is indicated below. While no new statistical issues arise, the additional information obtained can add biological insights and support or refute various conjectures about causes of mortality and causes of variation in mortality among beakers.

LeBlanc measured mortality on days 7, 14 and 21. Thus the overall 21 day mortality can be decomposed into 7 day mortality (early life mortality) and 21 day mortality conditional on survival for 7 days (later life mortality). This second measure of mortality eliminates the influences of the early mortality. Similarities or differences in the patterns across beakers and groups of these two components of mortality would tend to support or refute various conjectures about the reasons for the observed mortality. For example we observed strong dichotomies in 21 day mortality rates among beakers in group 6 of LeBlanc's Test A and among beakers in group 7 of LeBlanc's Test B. Did these dichotomies occur in early stage

mortality, in later stage mortality, or in both? To study conditional 21 day mortality given 7 day survival, we simply define mortality rates as 1 minus number live after 21 days divided by number live after 7 days. We then proceed as before. The only technical difference might be unequal "sample" sizes among beakers within groups or among groups. Of course, we could also condition on 14 day mortality.

We illustrate the use of such conditional measures of mortality by testing for beaker to beaker heterogeneity within groups in LeBlanc's Tests A and B using EXAX2.

LeBlanc Test A Day Mortality Conditional on 7 Day Survival

The results from the EXAX2 calculations are shown in Figures III.14 to III.20. They are summarized below.

LE BLANC TEST A 21 DAY MORTALITY CONDITIONAL ON 7 DAY SURVIVAL

Trt	Method	Chi sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	0.0000	1.0000	0.0000	1.66087	3.63660
2	exact	3.8095	0.6105	0.9869	0.8687	1.6419
3	exact	7.7714	0.0757	5.1619	1.5750	3.4893
4	exact	4.2270	0.3445	2.1311	1.5068	3.1226
5	exact	5.4795	0.1975	3.2436	1.5068	3.1226
6	exact	12.5778	0.0069	9.9396	1.9265	3.9809
7	exact	2.0000	1.0000	0.0000	0.0000	0.0000

$\Sigma\ -2\ell n\ A_i = 21.4632$ $\Sigma\ E(-2\ell n\ A_i) = 9.0447$ $\Sigma\ Var(-2\ell n\ A_i) = 18.9938$
$Z = 2.2433$ $P(Z > 2.2433) = 0.0124$

We can determine 7 day survival from the "TOTAL" column in Figures III.14 to III.20. Figures III.14 to III.18 show that there was no observed early mortality in groups 1-5. Figure III.19 shows a strong beaker to beaker dichotomy in 7 day mortality, very similar to that observed in the 21 day mortality. Figure III.20 shows almost complete mortality in group 7 by the seventh day. Thus the results in groups 1-5 and 7 with respect to conditional 21 day mortality are in complete agreement with those for unconditional 21 day mortality, namely no evidence of beaker to beaker heterogeneity, except perhaps a suggestion in group 3. The heterogeneity chi square in group 6 is very large and highly statistically significant, although no where near as large as that for the unconditional 21 day mortality. This is quite important because it tells us that the observed high mortality rates in beakers C and D and the relatively low rates in beakers A and B represent a persistent pattern throughout the entire duration of the test. Namely, of the seven survivors beyond day 7 in beaker C two died (i.e. 29 percent conditional mortality rate). Of the four survivors in beaker D three died (i.e. 75 percent conditional mortality rate). Beakers A and B had 10 percent and 6 percent conditional mortality rates respectively. Thus whatever caused

the dichotomy in mortality rates occurred either early in the test (thereby stressing the daphnids early) or else persistently throughout the test. Beaker D experienced by far the worst mortality rates, both in the early and the later stages of the test.

Overall there is evidence of significant beaker to beaker heterogeneity, due primarily to group 6. ($P[Z > 2.243] = 0.012$). We will study the magnitude of this variation in greater detail in later sections.

LeBlanc Test B 21 day Mortality Conditional on 7 Day Survival

The situation is similar to that for Test A. There is little 7 day mortality in any of the beakers in groups 1-6. Namely, the observed numbers of deaths at 7 days, by beaker, for each of these groups are (0, 1, 0, 0), (1, 2, 1, 1), (0, 1, 0, 2), (0, 0, 0, 0), (0, 0, 0, 0) and (0, 0, 0, 0). The conditional and unconditional 21 day mortality results should thus be similar in those groups. The results from the EXAX2 calculations are summarized below. Those for group 7 are shown in Figure III.21.

LE BLANC TEST B 21 DAY MORTALITY GIVEN 7 DAY SURVIVAL

Trt	Method	Chi sq	A_i	$-2\ell n\ A_i$	$E(-2\ell n\ A_i)$	$Var(-2\ell n\ A_i)$
1	exact	9.0690	0.0312	6.9371	1.8166	3.7415
2	exact	1.1096	0.8640	0.2925	1.8466	3.8394
3	exact	3.0928	0.4805	1.4658	1.0412	1.4213
4	exact	5.7600	0.1591	3.6768	1.3590	2.6405
5	exact	1.7451	0.7153	0.6702	1.7273	3.6654
6	exact	4.4444	0.2521	2.7562	1.5361	3.4424
7	exact	16.0178	0.0021	12.3604	1.8417	3.8074

$\Sigma\ -2\ell n\ A_i = 28.1589$ $\Sigma\ E(-2\ n\ A_i) = 11.1686$ $\Sigma\ Var(-2\ n\ A_i) = 22.5580$
$Z = 2.7647$ $P(Z > 2.7647) = 0.0028$

Figure III.21 shows a strong beaker to beaker dichotomy in 7 day mortality in group 7, with beakers A and D exhibiting substantially higher mortality than that in beakers B and C. Beakers A and D also exhibit substantially greater later stage mortality rates, especially beaker D. It is very interesting to note again that the beakers that exhibit the greater early stage mortality also exhibit the greatest later stage mortality. Thus the cause of the dichotomous mortality rates either occurred early in the test or else persisted throughout the test. These results are in direct correspondence with those from group 6 of Test A, thereby leading to a conjecture of a common cause for the dichtomous results in each group.

Overall there is evidence of statistically significant beaker to beaker heterogeneity, due primarily to group 7. We will study the magnitude of this variation in greater detail in later sections.

```
INVESTIGATOR: [illegible] -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (UG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ****************************************

DAPHNIA MAGNA MORTALITY

                    ****************************************

GROUP  1              CONTINGENCY TABLE

                     LIVE         DEAD          TOTAL
         TRT   1    17.000        3.000        20.000
         TRT   2    17.000        3.000        20.000
         TRT   3    17.000        3.000        20.000
         TRT   4    17.000        3.000        20.000

         TOTAL LIVE    =    68.000  , TOTAL DEAD    =    12.000  , GRAND TOTAL =    80.000

                 EXPECTED FREQUENCIES                              CUTOFF =   5.00000

                     LIVE         DEAD          TOTAL
         TRT   1    17.000        3.000        20.000
         TRT   2    17.000        3.000        20.000
         TRT   3    17.000        3.000        20.000
         TRT   4    17.000        3.000        20.000

         TOTAL LIVE    =    68.000  , TOTAL DEAD    =    12.000  , GRAND TOTAL =    80.000

          EXACT XSQ DISTRIBUTION
            XSQ         PROB       SIG                 14.90196     .00138     .00191
          0.00000      .02803    1.00000               17.25490     .00012     .00053
           .78431      .23829     .97197               18.82353     .00009     .00040
          1.56863      .06439     .73368               19.60784     .00025     .00031
          2.35294      .23369     .64928               21.17647     .00004     .00006
          3.13725      .08027     .41559               25.88235     .00001     .00002
          3.92157      .11371     .33532               26.66667     .00001     .00001
          4.70588      .07520     .22161               33.72549     .00000     .00000
          5.49020      .10100     .19641               42.35294     .00000     .00000
          6.27451      .00958     .09541
          7.05882      .03409     .08584         EXACT DISTRIBUTION OF XSQ STATISTIC USED.
          7.84314      .01421     .05174            OBSERVED XSQ =            0.00000
          8.62745      .01115     .03753            OBSERVED SIGNIFICANCE LEVEL = AI =  1.00000
          9.41176      .00704     .02638         YY = -2LN(AI) =          -.00000
         10.19608      .01148     .01934         EY = E(-2LN(AI)) =          1.66087
         11.76471      .00299     .00785         VARY = VAR(-2LN(AI)) =        3.63561
         13.33333      .00141     .00487
         14.11765      .00106     .00246
```

Figure III.1. EXAX2 output. LeBlanc Test A Group 1

```
INVESTIGATOR: [illegible] -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (AI/G) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)

*****************************************

DAPHNIA MAGNA MORTALITY

*****************************************

GROUP  2                CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
        TRT    1     18.000         2.000        20.000
        TRT    2     20.000         0.000        20.000
        TRT    3     20.000         0.000        20.000
        TRT    4     19.000         1.000        20.000

        TOTAL LIVE     =     77.000  , TOTAL DEAD     =      3.000   , GRAND TOTAL =     80.000

                EXPECTED FREQUENCIES                          CUTOFF =   5.00000

                    LIVE          DEAD          TOTAL
        TRT    1     19.250          .750        20.000
        TRT    2     19.250          .750        20.000
        TRT    3     19.250          .750        20.000
        TRT    4     19.250          .750        20.000

        TOTAL LIVE     =     77.000  , TOTAL DEAD     =      3.000   , GRAND TOTAL =     80.000

          EXACT XSQ DISTRIBUTION
            XSQ          PROB          SIG
          1.03896       .38948       1.00000
          3.40952       .55501        .61052
          9.45065       .05550        .05550

    EXACT DISTRIBUTION OF XSQ STATISTIC USED.
       OBSERVED XSQ =          3.40952
       OBSERVED SIGNIFICANCE LEVEL = AI =    .61052
    YY = -2LN(AI) =            .98190
    EY = E(-2LN(AI)) =          .86869
    VARY = VAR(-2LN(AI)) =     1.64159
```

Figure III.2. EXAX2 output. LeBlanc Test A Group 2

INVESTIGATOR: [illegible] -- TEST A

21 DAY FISH EMBRYO TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),

MEASURED CONC (0,7,14,21)

****

DAPHNIA MAGNA MORTALITY

****

GROUP 3 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	20.000	0.000	20.000
TRT	2	19.000	1.000	20.000
TRT	3	15.000	5.000	20.000
TRT	4	16.000	4.000	20.000

TOTAL LIVE = 70.000 , TOTAL DEAD = 10.000 , GRAND TOTAL = 80.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	17.500	2.500	20.000
TRT	2	17.500	2.500	20.000
TRT	3	17.500	2.500	20.000
TRT	4	17.500	2.500	20.000

TOTAL LIVE = 70.000 , TOTAL DEAD = 10.000 , GRAND TOTAL = 80.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.45714	.17097	1.00000
1.37143	.15272	.82903
2.28571	.30594	.67632
4.11429	.15169	.37038
5.02857	.04403	.20869
5.44286	.04405	.12465
7.77143	.04337	.07570
8.68571	.01320	.03234
9.60000	.01288	.02214
11.42857	.00988	.00926
12.34286	.00288	.00638
13.25714	.00429	.00350
15.08571	.00364	.00121
16.74286	.00057	.00057
17.65714	.00017	.00021
20.05714	.00002	.00002
21.28571	.00000	.00000

EXACT DISTRIBUTION OF XSQ STATISTIC USED,
OBSERVED XSQ = 7.77143
OBSERVED SIGNIFICANCE LEVEL = AI = .07570
YY = -2LN(AI) = 5.16186
EY = E(-2LN(AI)) = 1.57500
VARY = VAR(-2LN(AI)) = 3.48430

Figure III.3. EXAX2 output. LeBlanc Test A Group 3

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (UG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ****************************************
DAPHNIA MAGNA MORTALITY
                    ****************************************

GROUP   4            CONTINGENCY TABLE

                   LIVE          DEAD          TOTAL
        TRT   1   17.000         3.000        20.000
        TRT   2   20.000         0.000        20.000
        TRT   3   17.000         3.000        20.000
        TRT   4   19.000         1.000        20.000

        TOTAL LIVE   =    73.000  , TOTAL DEAD    =     7.000   , GRAND TOTAL =     80.000

                EXPECTED FREQUENCIES                            CUTOFF =   5.00000

                   LIVE          DEAD          TOTAL
        TRT   1   18.250         1.750        20.000
        TRT   2   18.250         1.750        20.000
        TRT   3   18.250         1.750        20.000
        TRT   4   18.250         1.750        20.000

        TOTAL LIVE   =    73.000  , TOTAL DEAD    =     7.000   , GRAND TOTAL =     80.000

          EXACT XSQ DISTRIBUTION
            XSQ          PROB        SIG
           .46967       .17273     1.00000
          1.72211       .32728      .82727
          2.97456       .15546      .49999
          4.22701       .14699      .34453
          5.47945       .13909      .19754
          7.98434       .02086      .05814
          9.23679       .02343      .03758
         10.48924       .01113      .01415
         15.49902       .00293      .00303
         23.01379       .00010      .00010

   EXACT DISTRIBUTION OF XSQ STATISTIC USED.
      OBSERVED XSQ =           4.22701
      OBSERVED SIGNIFICANCE LEVEL = AI =     .34453
   YY = -2LN(AI) =           2.13116
   FY = E(-2LN(AI)) =            1.50684
   VARY = VAR(-2LN(AI)) =            3.12255
```

Figure III.4. EXAX2 output. LeBlanc Test A Group 4

```
INVESTIGATOR: LEBLANC -- TEST A
21 DAY FLOW LOW LEVEL TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) ID, TESTED, NO, LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ***** ***********************************
DAPHNIA MAGNA MORTALITY
                    *****************************************

GROUP   5             CONTINGENCY TABLE

                     LIVE         DEAD          TOTAL
          TRT   1    18.000       2.000         20.000
          TRT   2    20.000       0.000         20.000
          TRT   3    16.000       4.000         20.000
          TRT   4    19.000       1.000         20.000

          TOTAL LIVE    =    73.000  , TOTAL DEAD    =    7.000  , GRAND TOTAL =    80.000

                  EXPECTED FREQUENCIES                          CUTOFF =   5.00000

                     LIVE         DEAD          TOTAL
          TRT   1    18.250       1.750         20.000
          TRT   2    18.250       1.750         20.000
          TRT   3    18.250       1.750         20.000
          TRT   4    18.250       1.750         20.000

          TOTAL LIVE    =    73.000  , TOTAL DEAD    =    7.000  , GRAND TOTAL =    80.000

           EXACT XSQ DISTRIBUTION
            XSQ          PROB       SIG
            .46967       .17273     1.00000
           1.72211       .32728      .82727
           2.47456       .15546      .49999
           4.22701       .14699      .34453
           5.47945       .13909      .19754
           7.98431       .02186      .05844
           9.23679       .02343      .03758
          10.48924       .01113      .01415
          15.49902       .00243      .00303
          23.01370       .00010      .00010

      EXACT DISTRIBUTION OF XSQ STATISTIC USED.
         OBSERVED XSQ =             5.47945
         OBSERVED SIGNIFICANCE LEVEL = AI =     .19754
      YY = -2LN(AI) =             3.24364
      EY = E(-2LN(AI)) =             1.56844
      VARY = VAR(-2LN(AI)) =            3.12255
```

Figure III.5. EXAX2 output. LeBlanc Test A Group 5

```
INVESTIGATOR: E. A. G. BLANCS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ******************************************

DAPHNIA MAGNA MORTALITY
                    ******************************************

GROUP   6                CONTINGENCY TABLE

                      LIVE          DEAD          TOTAL
        TRT     1     18.000         2.000        20.000
        TRT     2     16.000         4.000        20.000
        TRT     3      5.000        15.000        20.000
        TRT     4      1.000        19.000        20.000

        TOTAL LIVE     =     40.000   , TOTAL DEAD     =     40.000   , GRAND TOTAL =     80.000

                  EXPECTED FREQUENCIES                          CUTOFF =   5.00000

                      LIVE          DEAD          TOTAL
        TRT     1     10.000        10.000        20.000
        TRT     2     10.000        10.000        20.000
        TRT     3     10.000        10.000        20.000
        TRT     4     10.000        10.000        20.000

        TOTAL LIVE     =     40.000   , TOTAL DEAD     =     40.000   , GRAND TOTAL =     80.000

   ASYMPTOTIC CHI SQUARED USED.
      OBSERVED CHI SQUARED =          31.27500
      OBSERVED SIGNIFICANCE LEVEL = AI =    .00000
   YY = -2LN(AI) =          37.88625
   EY = E(-2LN(AI)) =        2.00000
   VARY = VAR(-2LN(AI)) =          4.00000
```

Figure III.6. EXAX2 output. LeBlanc Test A Group 6

```
INVESTIGATOR: [illegible] -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRT GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ****************************************

DAPHNIA MAGNA MORTALITY

                    ****************************************

GROUP  7              CONTINGENCY TABLE

                      LIVE          DEAD          TOTAL
          TRT    1     0.000        20.000        20.000
          TRT    2     0.000        20.000        20.000
          TRT    3     1.000        19.000        20.000
          TRT    4     0.000        20.000        20.000

          TOTAL LIVE  =     1.000  , TOTAL DEAD  =    79.000  , GRAND TOTAL =    80.000

                EXPECTED FREQUENCIES                          CUTOFF =   5.00000

                      LIVE          DEAD          TOTAL
          TRT    1      .250        19.750        20.000
          TRT    2      .250        19.750        20.000
          TRT    3      .250        19.750        20.000
          TRT    4      .250        19.750        20.000

          TOTAL LIVE  =     1.000  , TOTAL DEAD  =    79.000  , GRAND TOTAL =    80.000

            EXACT XSQ DISTRIBUTION
             XSQ          PROB         SIG
            3.03797     1.00000     1.00000

      EXACT DISTRIBUTION OF XSQ STATISTIC USED.
         OBSERVED XSQ =          3.03797
         OBSERVED SIGNIFICANCE LEVEL = AI =  1.00000
      YY = -2LN(AI) =             .00000
      EY = E(-2LN(AI)) =          .00000
      VARY = VAR(-2LN(AI)) =      .00000

    7   INDEPENDENT TESTS COMBINED
      SY = SUM(YY) =           49.40982
      YMJ = SUM(EY) =           9.11825
      SVARY = SUM(VARY) =      19.01290

      Z = (SUM(SY) - SQRT(YS(I))/SQRT(SVARYY/(4*YMJ)) =     5.55340
      UNDER HO Z HAS A STANDARD NORMAL DISTRIBUTION
```

Figure III.7. EXAX2 output. LeBlanc Test A Group 7 and combination of results across groups.

GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

****

DAPHNIA MAGNA MORTALITY

****

GROUP 1 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	7.000	0.000	7.000
TRT	2	13.000	2.000	15.000

TOTAL LIVE = 20.000 , TOTAL DEAD = 2.000 , GRAND TOTAL = 22.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	6.364	.636	7.000
TRT	2	13.636	1.364	15.000

TOTAL LIVE = 20.000 , TOTAL DEAD = 2.000 , GRAND TOTAL = 22.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.33524	.45455	1.00000
1.02667	.45455	.54545
4.71429	.09091	.09091

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
OBSERVED XSQ = 1.02667
OBSERVED SIGNIFICANCE LEVEL = AI = .54545

YY = -2LN(AI) = 1.21227
EY = E(-2LN(AI)) = .98701
VARY = VAR(-2LN(AI)) = 1.78468

Figure III.8. EXAX2 output. Goulden - Isophorone. Group 1. Comparison of mortality between individually and multiply housed daphnids.

GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

****

DAPHNIA MAGNA MORTALITY

****

GROUP 2 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	6.000	1.000	7.000
TRT	2	14.000	1.000	15.000

TOTAL LIVE = 20.000 , TOTAL DEAD = 2.000 , GRAND TOTAL = 22.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	6.364	.636	7.000
TRT	2	13.636	1.364	15.000

TOTAL LIVE = 20.000 , TOTAL DEAD = 2.000 , GRAND TOTAL = 22.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.33524	.45455	1.00000
1.02667	.45455	.54545
4.71429	.09091	.09091

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
OBSERVED XSQ = .33524
OBSERVED SIGNIFICANCE LEVEL = AI = 1.00000
YY = -2LN(AI) = .00000
EY = E(-2LN(AI)) = .98701
VARY = VAR(-2LN(AI)) = 1.78468

Figure III.9. EXAX2 output. Goulden - Isophorone. Group 2. Comparison of mortality between individually and multiply housed daphnids.

```
GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

*****************************************

DAPHNIA MAGNA MORTALITY

*****************************************

GROUP   3              CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
        TRT    1     7.000         0.000         7.000
        TRT    2    14.000         1.000        15.000

        TOTAL LIVE  =    21.000  , TOTAL DEAD  =     1.000  , GRAND TOTAL =    22.000

              EXPECTED FREQUENCIES                          CUTOFF =   5.00000

                    LIVE          DEAD          TOTAL
        TRT    1     6.682          .318         7.000
        TRT    2    14.318          .682        15.000

        TOTAL LIVE  =    21.000  , TOTAL DEAD  =     1.000  , GRAND TOTAL =    22.000

          EXACT XSQ DISTRIBUTION
           XSQ          PROB        SIG
           .48889       .68182     1.00000
          2.24490       .31818      .31818

    EXACT DISTRIBUTION OF XSQ STATISTIC USED.
       OBSERVED XSQ =            .48889
       OBSERVED SIGNIFICANCE LEVEL = AI =  1.00000
    YY = -2LN(AI) =              .00000
    EY = E(-2LN(AI)) =           .72872
    VARY = VAR(-2LN(AI)) =      1.13793
```

Figure III.10. EXAX2 output. Goulden - Isophorone. Group 3. Comparison of mortality between individually and multiply housed daphnids.

GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

****

DAPHNIA MAGNA MORTALITY

****

GROUP 4 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	7.000	0.000	7.000
TRT	2	12.000	3.000	15.000

TOTAL LIVE = 19.000 , TOTAL DEAD = 3.000 , GRAND TOTAL = 22.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	6.045	.955	7.000
TRT	2	12.955	2.045	15.000

TOTAL LIVE = 19.000 , TOTAL DEAD = 3.000 , GRAND TOTAL = 22.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.00368	.47727	1.00000
1.62105	.29545	.52273
1.94453	.20455	.22727
7.44361	.02273	.02273

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
OBSERVED XSQ = 1.62105
OBSERVED SIGNIFICANCE LEVEL = AI = .52273
YY = -2LN(AI) = 1.29739
EY = E(-2LN(AI)) = 1.16144
VARY = VAR(-2LN(AI)) = 2.24623

Figure III.11. EXAX2 output. Goulden - Isophorone. Group 4. Comparison of mortality between individually and multiply housed daphnids.

GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

****

DAPHNIA MAGNA MORTALITY

****

GROUP 5 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	7.000	0.000	7.000
TRT	2	7.000	8.000	15.000

TOTAL LIVE = 14.000 , TOTAL DEAD = 8.000 , GRAND TOTAL = 22.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	4.455	2.545	7.000
TRT	2	9.545	5.455	15.000

TOTAL LIVE = 14.000 , TOTAL DEAD = 8.000 , GRAND TOTAL = 22.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.18707	.32869	1.00000
.26939	.32869	.67131
1.91565	.14940	.34262
2.16259	.14087	.19322
5.45510	.02988	.05235
5.86667	.02012	.02247
10.80544	.00230	.00235
17.96667	.00005	.00005

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
OBSERVED XSQ = 5.86667
OBSERVED SIGNIFICANCE LEVEL = AI = .02247
YY = -2LN(AI) = 7.59121
EY = E(-2LN(AI)) = 1.40302
VARY = VAR(-2LN(AI)) = 3.00424

Figure III.12. EXAX2 output. Goulden - Isophorone. Group 5. Comparison of mortality between individually and multiply housed daphnids.

```
GOULDEN SURVIVAL DATA--REPS 1-7(POOLED) VS. REPS 8-10(POOLED)

*****************************************

DAPHNIA MAGNA MORTALITY

*****************************************

GROUP   6              CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
        TRT    1      1.000         6.000         7.000
        TRT    2      2.000        13.000        15.000

        TOTAL LIVE     =     3.000  , TOTAL DEAD     =     19.000   , GRAND TOTAL =     22.000

              EXPECTED FREQUENCIES                                   CUTOFF =   5.00000

                    LIVE          DEAD          TOTAL
        TRT    1       .955        6.045         7.000
        TRT    2      2.045       12.955        15.000

        TOTAL LIVE     =     3.000  , TOTAL DEAD     =     19.000   , GRAND TOTAL =     22.000

          EXACT XSQ DISTRIBUTION
           XSQ          PROB        SIG
           .00368       .47727     1.00000
          1.62105       .29545      .52273
          1.94453       .20455      .22727
          7.44361       .02273      .02273

    EXACT DISTRIBUTION OF XSQ STATISTIC USED.
       OBSERVED XSQ =          .00368
       OBSERVED SIGNIFICANCE LEVEL = AI =  1.00000
    YY = -2LN(AI) =          .00000
    EY = E(-2LN(AI)) =       1.16144
    VARY = VAR(-2LN(AI)) =   2.24623

  6   INDEPENDENT TESTS COMBINED
    SY = SUM(YY) =          10.10088
    YMU = SUM(EY) =     6.42864
    SVARY = SUM(VARY) =    12.20399

    Z = (SQRT(SY) - SQRT(YMU))/SQRT(SVARY/(4*YMU)) =      .93294
    UNDER HO Z HAS A STANDARD NORMAL DISTRIBUTION
```

80

Figure III.13. EXAX2 output. Goulden - Isophorone. Group 6. Comparison of mortality between individually and multiply housed daphnids. Combination of results across groups.

```
        14.90196      .00138      .00191
        17.25490      .00012      .00053
        18.82353      .00009      .00040
        19.50784      .00025      .00031
        21.17647      .00004      .00006
        25.88235      .00001      .00002
        26.66667      .00001      .00001
        33.72549      .00000      .00000
        42.35294      .00000      .00000

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
   OBSERVED XSQ =          0.00000
   OBSERVED SIGNIFICANCE LEVEL = AI =  1.00000
YY = -2LN(AI) =           -.00000
EY = E(-2LN(AI)) =         1.66087
VARY = VAR(-2LN(AI)) =     3.63660
```


Figure III.14. EXAX2 output. LeBlanc Test A. Group 1. 21 day mortality conditional on 7 day survival.

INVESTIGATION: EC AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

****

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

****

GROUP 1 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	17.000	3.000	20.000
TRT	2	17.000	3.000	20.000
TRT	3	17.000	3.000	20.000
TRT	4	17.000	3.000	20.000

TOTAL LIVE = 68.000 , TOTAL DEAD = 12.000 , GRAND TOTAL = 80.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	17.000	3.000	20.000
TRT	2	17.000	3.000	20.000
TRT	3	17.000	3.000	20.000
TRT	4	17.000	3.000	20.000

TOTAL LIVE = 68.000 , TOTAL DEAD = 12.000 , GRAND TOTAL = 80.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
0.00000	.02803	1.00000
.78431	.23829	.97197
1.56863	.08439	.73368
2.35294	.23369	.64928
3.13725	.08027	.41559
3.92157	.11371	.33532
4.70588	.02520	.22161
5.49020	.10100	.19641
6.27451	.00958	.09541
7.05882	.03409	.08584
7.84314	.01421	.05174
8.62745	.01115	.03753
9.41176	.00704	.02638
10.19608	.01148	.01934
11.76471	.00299	.00786
13.33333	.00191	.00487
14.11765	.00106	.00295

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ****************************************
21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.
                    ****************************************

GROUP  2                CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
      TRT    1      18.000         2.000        20.000
      TRT    2      20.000         0.000        20.000
      TRT    3      20.000         0.000        20.000
      TRT    4      19.000         1.000        20.000

      TOTAL LIVE  =     77.000  , TOTAL DEAD  =      3.000  , GRAND TOTAL =     80.000

                EXPECTED FREQUENCIES                          CUTOFF =  5.00000

                    LIVE          DEAD          TOTAL
      TRT    1      19.250          .750        20.000
      TRT    2      19.250          .750        20.000
      TRT    3      19.250          .750        20.000
      TRT    4      19.250          .750        20.000

      TOTAL LIVE  =     77.000  , TOTAL DEAD  =      3.000  , GRAND TOTAL =     80.000

        EXACT XSQ DISTRIBUTION
          XSQ          PROB         SIG
        1.03896        .38948      1.00000
        3.80952        .55501       .61052
        9.35065        .05550       .05550

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
   OBSERVED XSQ =          3.80952
   OBSERVED SIGNIFICANCE LEVEL = AI =    .61052
YY = -2LN(AI) =          .98690
EY = E(-2LN(AI)) =         .86869
VARY = VAR(-2LN(AI)) =     1.64189
```

Figure III.15. EXAX2 output. LeBlanc Test A. Group 2. 21 day mortality conditional on 7 day survival.

INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),

MEASURED CONC (0,7,14,21)

****

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

****

GROUP 3 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	20.000	0.000	20.000
TRT	2	19.000	1.000	20.000
TRT	3	15.000	5.000	20.000
TRT	4	16.000	4.000	20.000

TOTAL LIVE = 70.000 , TOTAL DEAD = 10.000 , GRAND TOTAL = 80.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	17.500	2.500	20.000
TRT	2	17.500	2.500	20.000
TRT	3	17.500	2.500	20.000
TRT	4	17.500	2.500	20.000

TOTAL LIVE = 70.000 , TOTAL DEAD = 10.000 , GRAND TOTAL = 80.000

EXACT XSJ DISTRIBUTION

XSJ	PROB	SIG
.45714	.17097	1.00000
1.37143	.15272	.82903
2.28571	.30594	.67632
4.11429	.16169	.37038
5.02857	.09403	.20869
5.94286	.04895	.12465
7.77143	.04337	.07570
8.68571	.01020	.03234
9.60000	.01288	.02214
11.42857	.00088	.00926
12.34286	.00288	.00838
13.25714	.00429	.00550
15.08571	.00064	.00121
18.74286	.00037	.00057
19.65714	.00017	.00020
26.05714	.00002	.00002
33.28571	.00000	.00000

```
EXACT DISTRIBUTION OF XSQ STATISTIC USED.
   OBSERVED XSQ =          7.77143
   OBSERVED SIGNIFICANCE LEVEL = AI =    .07570
YY = -2LN(AI) =          5.16186
EY = E(-2LN(AI)) =          1.57500
VARY = VAR(-2LN(AI)) =          3.48930
```


Figure III.16. EXAX2 output. LeBlanc Test A. Group 3. 21 day mortality conditional on 7 day survival.

```
INVESTIGATOR: FG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ******************************************

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

                    ******************************************

GROUP  4                CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
          TRT   1     17.000        3.000        20.000
          TRT   2     20.000        0.000        20.000
          TRT   3     17.000        3.000        20.000
          TRT   4     19.000        1.000        20.000

          TOTAL LIVE  =     73.000  , TOTAL DEAD     =      7.000   , GRAND TOTAL =     80.000

                    EXPECTED FREQUENCIES                          CUTOFF =  5.00000

                    LIVE          DEAD          TOTAL
          TRT   1     18.250        1.750        20.000
          TRT   2     18.250        1.750        20.000
          TRT   3     18.250        1.750        20.000
          TRT   4     18.250        1.750        20.000

          TOTAL LIVE  =     73.000  , TOTAL DEAD     =      7.000   , GRAND TOTAL =     80.000

          EXACT XSQ DISTRIBUTION
            XSQ          PROB          SIG
            .46967       .17273       1.00000
           1.72211       .32728        .82727
           2.97456       .15546        .49939
           4.22701       .14699        .34453
           5.47945       .13909        .19754
           7.98434       .07086        .05844
           9.23679       .02343        .03758
          10.48924       .01113        .01415
          15.49902       .00293        .00303
          23.01370       .00010        .00010

     EXACT DISTRIBUTION OF XSQ STATISTIC USED.
        OBSERVED XSQ =          4.22701
        OBSERVED SIGNIFICANCE LEVEL = AI =    .34453
     YY = -2LN(AI) =          2.13116
     EY = E(-2LN(AI)) =       1.50684
     VARY = VAR(-2LN(AI)) =   4.12254
```

Figure III.17. EXAX2 output. LeBlanc Test A. Group 4. 21 day mortality conditional on 7 day survival

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),

MEASURED CONC (0,7,14,21)

*****************************************

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

*****************************************

GROUP  5              CONTINGENCY TABLE

                  LIVE          DEAD          TOTAL
     TRT    1     18.000         2.000        20.000
     TRT    2     20.000         0.000        20.000
     TRT    3     15.000         4.000        20.000
     TRT    4     19.000         1.000        20.000

     TOTAL LIVE     =     73.000  , TOTAL DEAD     =      7.000   , GRAND TOTAL =     80.000

                EXPECTED FREQUENCIES                          CUTOFF =   5.00500

                  LIVE          DEAD          TOTAL
     TRT    1     18.250         1.750        20.000
     TRT    2     18.250         1.750        20.000
     TRT    3     18.250         1.750        20.000
     TRT    4     18.250         1.750        20.000

     TOTAL LIVE     =     73.000  , TOTAL DEAD     =      7.000   , GRAND TOTAL =     80.000

  EXACT XSQ DISTRIBUTION
       XSQ          PROB          SIG
     .46967        .17273       1.00000
    1.72211        .32728        .82727
    2.97456        .15546        .49999
    4.22701        .14599        .34453
    5.47945        .13909        .19754
    7.98434        .02086        .05844
    9.23679        .02343        .03758
   10.48924        .01113        .01415
   15.49902        .00293        .00303
   23.01370        .00010        .00010

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
   OBSERVED XSQ =          5.47945
   OBSERVED SIGNIFICANCE LEVEL = AI =    .19754
YY = -2LN(AI) =          3.24364
EY = E(-2LN(AI)) =          1.50684
VARY = VAR(-2LN(AI)) =          3.12755
```

Figure III.18. EXAX2 output. LeBlanc test A. Group 5 21 day mortality conditional on 7 day survival

INVESTIGATOR: FG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),

MEASURED CONC (0,7,14,21)

****

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

****

GROUP 6 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	18.000	2.000	20.000
TRT	2	16.000	1.000	17.000
TRT	3	5.000	2.000	7.000
TRT	4	1.000	3.000	4.000

TOTAL LIVE = 40.000 , TOTAL DEAD = 8.000 , GRAND TOTAL = 48.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	16.667	3.333	20.000
TRT	2	14.167	2.833	17.000
TRT	3	5.833	1.167	7.000
TRT	4	3.333	.667	4.000

TOTAL LIVE = 40.000 , TOTAL DEAD = 8.000 , GRAND TOTAL = 48.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.28034	.05752	1.00000
.68269	.04889	.94248
1.00034	.06112	.89359
1.24840	.03451	.83247
1.44504	.08389	.79796
1.56605	.07190	.71407
1.77176	.03492	.64217
1.96840	.03567	.60725
2.12269	.03911	.57058
2.21647	.02876	.53146
2.49782	.01833	.50270
2.65213	.01956	.48437
2.73076	.02517	.46481
2.89412	.02235	.43964
2.98647	.03056	.41729
3.21176	.02794	.38673
3.45076	.01960	.35880

Groups (Cont.)

3.45681	.02181	.34820
3.56269	.01726	.32639
3.88036	.01438	.30913
3.93782	.01467	.29475
4.17681	.00918	.28008
4.22824	.03116	.27090
4.59126	.02397	.23974
4.65176	.01233	.21578
4.81210	.02139	.20345
4.84840	.00863	.18206
4.90891	.01198	.17343
5.15412	.02445	.16145
5.12067	.00719	.13700
5.31429	.00345	.12981
5.37782	.00647	.12636
5.46252	.00589	.11989
5.58353	.00698	.11330
5.62891	.00505	.10602
5.76504	.00530	.10097
5.81647	.00719	.09567
5.84067	.00764	.08848
5.88605	.00454	.08084
6.18857	.00288	.07630
6.79361	.00442	.07343
6.90252	.00138	.06901
7.02353	.00419	.06763
7.03563	.00459	.06344
7.08706	.00523	.05885
7.21714	.00180	.05262
7.47429	.00377	.05082
7.62857	.00173	.04705
7.80706	.00262	.04532
8.23361	.00098	.04270
8.46353	.00349	.04181
8.47563	.00092	.03832
8.49076	.00079	.03740
8.54824	.00197	.03661
8.65714	.00144	.03464
8.72067	.00180	.03320
8.91126	.00151	.03141
9.06857	.00144	.02989
9.21681	.00069	.02845
9.76134	.00069	.02777
9.84000	.00082	.02708
9.99126	.00240	.02626
10.09714	.00063	.02386
10.24538	.00057	.02322
10.52067	.00300	.02265
10.56000	.00062	.01965
10.66891	.00038	.01904
10.71126	.00101	.01866
10.76269	.00192	.01765
11.02891	.00126	.01573
11.29210	.00144	.01447
11.51294	.00103	.01303
11.53412	.00164	.01200
11.60034	.00101	.01036
11.85176	.00137	.00935
11.89714	.00042	.00798
12.18151	.00046	.00756
12.46891	.00025	.00720
12.57782	.00072	.00634
12.61714	.00035	.00624

12.76840	.00361	.00578
12.78353	.00097	.00517
12.84706	.00320	.00433
12.95294	.00021	.00419
13.38857	.00036	.00389
13.44000	.00033	.00354
13.63361	.00022	.00320
13.73647	.00050	.00298
13.95429	.00025	.00248
14.40504	.00018	.00222
14.52605	.00015	.00205
14.84067	.00019	.00189
15.49714	.00011	.00170
15.60000	.00016	.00160
15.75126	.00008	.00143
16.64067	.00013	.00136
17.18824	.00007	.00123
17.29714	.00007	.00117
17.50588	.00006	.00110
17.55126	.00005	.00103
17.97782	.00018	.00098
18.16840	.00015	.00080
19.69782	.00008	.00065
19.35429	.00011	.00058
19.60840	.00003	.00047
19.92605	.00004	.00044
22.09210	.00006	.00040
22.28269	.00005	.00034
22.33412	.00007	.00029
22.86353	.00005	.00022
23.37176	.00004	.00017
23.41714	.00001	.00014
23.46857	.00002	.00013
23.48067	.00001	.00011
23.67429	.00001	.00010
23.73782	.00001	.00009
24.04034	.00001	.00009
24.09782	.00002	.00008
24.75429	.00001	.00006
24.96000	.00001	.00005
25.00840	.00001	.00003
25.97647	.00001	.00003
28.21210	.00001	.00002
28.81714	.00000	.00001
28.86857	.00000	.00001
28.98067	.00000	.00001
29.12269	.00000	.00001
29.58857	.00000	.00000
29.65210	.00000	.00000
32.31429	.00000	.00000
34.62857	.00000	.00000
35.65714	.00000	.00000
41.16000	.00000	.00000
41.22353	.00000	.00000
42.60000	.00000	.00000

EXACT DISTRIBUTION OF KSD STATISTIC USED.
OBSERVED KSD = 12.57782
OBSERVED SIGNIFICANCE LEVEL = AI = .00594
YY = -2LN(AI) = 9.93962
EY = E(-2LN(AI)) = 1.92645
VARY = VAR(-2LN(AI)) = 3.98086

Figure III.19. EXAX2 output. LeBlanc test A. Group 6.
21 day mortality conditional on 7 day survival

INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

****

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

****

GROUP 7 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	0.000	0.000	0.000
TRT	2	0.000	0.000	0.000
TRT	3	1.000	0.000	1.000
TRT	4	0.000	1.000	1.000

TOTAL LIVE = 1.000 , TOTAL DEAD = 1.000 , GRAND TOTAL = 2.000

TABLE HAS BEEN REDUCED TO 2 COLUMNS

		LIVE	DEAD	TOTAL
TRT	1	1.000	0.000	1.000
TRT	2	0.000	1.000	1.000

TOTAL LIVE = 1.000 , TOTAL DEAD = 1.000 , GRAND TOTAL = 2.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	.500	.500	1.000
TRT	2	.500	.500	1.000

TOTAL LIVE = 1.000 , TOTAL DEAD = 1.000 , GRAND TOTAL = 2.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
2.00000	1.00000	1.00000

EXACT DISTRIBUTION OF XSQ STATISTIC USED.
OBSERVED XSQ = 2.00000
OBSERVED SIGNIFICANCE LEVEL = AI = 1.00000
YY = -2LN(AI) = .00000
EY = E(-2LN(AI)) = .00000
VARY = VAR(-2LN(AI)) = 0.00000

7 INDEPENDENT TESTS OBTAINED

```
SY = SUM(YY) =          21.46319
YMU = SUM(SY) =      9.04470
SVARY = SUM(VARY) =    18.99376

Z = (SQRT(SY) - SQRT(YMU))/SQRT(SVARY/(4*YMU)) =     2.24327
UNDER HO Z HAS A STANDARD NORMAL DISTRIBUTION
```


Figure III.20. EXAX2 output. LeBlanc test A. Group 7.
21 day mortality conditional on 7 day survival.

INVESTIGATION: FG AND G BIONOMICS--TEST B

21 DATA FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP - FORMAT: (I1,1X,A1,2F7.4,4I5,3F5.1,4(A1,F7.5)

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21),CUM. PERT. (7,14,21,),

MEASURED CONC (0,7,14,21)

****

21-DAY MORTALITY GIVEN SURVIVAL TO DAY 7, EXACT DISTRIB.

****

GROUP 7 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	3.000	2.000	5.000
TRT	2	14.000	0.000	14.000
TRT	3	12.000	0.000	12.000
TRT	4	1.000	2.000	3.000

TOTAL LIVE = 30.000 , TOTAL DEAD = 4.000 , GRAND TOTAL = 34.000

EXPECTED FREQUENCIES CUTOFF = 5.00000

		LIVE	DEAD	TOTAL
TRT	1	4.412	.588	5.000
TRT	2	12.353	1.647	14.000
TRT	3	10.588	1.412	12.000
TRT	4	2.647	.353	3.000

TOTAL LIVE = 30.000 , TOTAL DEAD = 4.000 , GRAND TOTAL = 34.000

EXACT XSQ DISTRIBUTION

XSQ	PROB	SIG
.94849	.11773	1.00000
1.29254	.09962	.88227
1.43016	.17951	.78265
2.09532	.05434	.65314
2.23294	.07064	.59880
2.66239	.09419	.52816
2.57698	.05977	.43397
3.35683	.02943	.37420
3.37976	.06641	.34477
3.58619	.03924	.27836
3.81556	.02135	.23911
4.51833	.02372	.21776
4.66421	.03623	.19404
4.87063	.02355	.15782
5.30278	.01423	.13427
5.92571	.01962	.12004
6.38344	.01423	.10042

```
 6.47619      .02158      .08619
 7.07254      .00906      .06460
 7.18722      .00776      .05555
 8.41111      .01067      .04778
 9.80198      .01087      .03711
10.92587      .00453      .02624
11.04056      .00388      .02171
11.06349      .00589      .01783
11.52222      .00427      .01195
13.49476      .00302      .00768
13.60944      .00259      .00466
16.01778      .00129      .00207
25.05476      .00030      .00078
25.16944      .00026      .00047
26.29333      .00022      .00022

EXACT DISTRIBUTION OF XSQ STATISTIC JSED.
   OBSERVED XSQ =          16.01778
   OBSERVED SIGNIFICANCE LEVEL = AI =    .00207
YY = -2LN(AI) =          12.36038
EY = E(-2LN(AI)) =        1.84166
VARY = VAR(-2LN(AI)) =    3.80742

7   INDEPENDENT TESTS COMBINED
SY = SUM(YY) =           28.15891
YMU = SUM(EY) =          11.16860
SVARY = SUM(VARY) =      22.55800

Z = (SQRT(SY) - SQRT(YMU))/SQRT(SVARY/(4*YMU)) =    2.76466
UNDER H0 Z HAS A STANDARD NORMAL DISTRIBUTION
```

Figure III.21. EXAX 2 output. LeBlanc test B. Group 7. 21 day mortality conditional on 7 day survival. Combination of results across groups.

IV. TESTING FOR BEAKER TO BEAKER HETEROGENEITY WITHIN TREATMENT GROUPS - LENGTH DATA

A. BACKGROUND

In the previous section we discussed the comparison of mortality rates across beakers within groups to test for heterogeneity of responses. Similar comparisons might be made for the production and length responses. There is difficulty however with respect to comparisons of production measurements. Since production is determined and reported on a beaker basis, there is just one determination per beaker and no internal estimate of its variability. In order to test for beaker to beaker heterogeneity we either need internal estimates of variability among daphnids within beakers, such as we have for the length data, or else we need external estimates of variability based on a theoretical model, such as we have for the survival data (i.e. binomial model). Thus unless we can hypothesize a theoretical statistical model that should govern production as a function of dose and time, we have no alternative but to carry out further statistical analyses on a beaker basis.

The situation is different with respect to length data. Lengths are measured and reported on a per daphnid basis. We can thus compare average lengths across beakers by analysis of variance techniques, using variability among daphnids within beakers as an error yardstick. We carry out such comparisons below for LeBlanc's Tests A and B. Adams and Goulden report no length data. In Chapman's test there is just one daphnid per beaker, so there is nothing to compare.

B. ANALYSIS OF VARIANCE TESTS FOR HETEROGENEITY AMONG BEAKERS WITHIN GROUPS IN LE BLANC'S TESTS A AND B

In this subsection we illustrate the comparisons of average 21 day lengths among replicate beakers within groups.

LeBlanc Test A 21 Day Lengths

There are a water control group, a solvent control group, and five treatment groups. Group 7, the highest treatment group, had just one survivor. There are thus no comparisons to be made. Comparisons among beakers within groups can be made by fitting a two way nested analysis of variance model to the data. Alternatively one way analysis of variance models can be fitted separately to the responses within groups and the results pooled across groups. This was done for the data from groups 1-6. The results of these calculations are shown in Figures IV.1 to IV.7 and are summarized below. The average lengths within beakers are displayed graphically in Figure II.26 and the standard deviations are displayed in Figure II.27. The standard deviation plot is not suggestive of any particular standard transformation to be carried out on the lengths before analysis. We carry out comparisons on the untransformed lengths.

LE BLANC TEST A 21 DAY LENGTHS

Trt	Between Beakers SS	d.f.	Within Beakers SS	d.f.	Within Group Significance Level
1	1.2335	3	8.1753	64	0.029
2	0.1941	3	7.0090	73	0.571
3	1.9310	3	8.3227	66	0.003
4	0.5636	3	4.8759	69	0.055
5	0.8066	3	8.3178	69	0.092
6	0.1708	3	5.4292	36	0.770
	4.8996	18	42.1299	377	

ANOVA TABLE (NESTED)

Source	d.f.	Sum of Squares	Mean Square
Between Groups	5	5.5815	1.1163
Between Beakers Within Groups	18	4.8996	0.2722
Within Beakers	377	42.1229	0.1118
Total	400	52.6110	

We test for significant beaker to beaker variation within groups by comparing the between beakers within groups mean square to the within beaker mean square. This ratio has an F distribution with 18 and 377 d.f. under the null hypothesis of no beaker to beaker variation.

$$F = \frac{0.2722}{0.1118} = 2.435. \qquad \text{Significant at } \alpha = 0.001$$

There is thus strong statistical evidence of beaker to beaker variation in lengths within groups. This is not a property of just one group or just one beaker since four of the six groups show significant F ratios based on just the data within these groups.

LeBlanc Test B 21 Day Lengths

The layout of the test is similar to that for Test A. Since treatment group 7 has more survivors in Test B than in Test A, we include it in the comparisons here. As before, one way analysis of variance models were fitted separately to the responses within groups and the results pooled across groups to yield a nested two way analysis of variance. The results of these

calculations are shown in Figures IV.8 to IV.15 and are summarized below. The average lengths within beakers are displayed graphically in Figure II.29 and the standard deviations are displayed in Figure II.30. The standard deviations do not seem to vary with length and we again perform no transformations on the length.

LE BLANC TEST B 21 DAY LENGTHS

Trt	Between Beakers SS	d.f.	Within Beakers SS	d.f.	Within Groups Significant Level
1	0.1633	3	6.5761	68	0.641
2	1.6418	3	7.5000	63	0.006
3	1.6130	3	14.8211	72	0.058
4	0.3889	3	9.7439	71	0.424
5	0.1878	3	6.4462	63	0.610
6	1.4236	3	11.0764	68	0.041
7	0.266	3	3.6130	26	0.598
	5.6844	21	59.7767	431	

ANOVA TABLE (NESTED)

Source	d.f.	Sum of Squares	Mean Square
Between Groups	6	1.7490	0.2915
Between Beakers Within Groups	21	5.6844	0.2707
Within Beakers	431	59.7767	0.1387
Total	458	67.2098	

We test for significant beaker to beaker variation within groups by comparing the between beakers within groups mean square to the within beakers mean square. This ratio has an F distribution with 21 and 431 d.f. under the null hypothesis of no beaker to beaker variation.

$$F = \frac{0.2707}{0.1387} = 1.952. \quad \text{Significant at } \alpha = 0.006$$

There is thus strong statistical evidence of beaker to beaker variation in lengths within groups. This result is in direct agreement with that for Test A.

In summary, we have evidence of beaker to beaker variation in lengths in both Tests A and B. Subsequent analyses will have to account for this.

TREATMENT GROUP AVERAGES
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP1 GRP2 GRP3 GRP4 GRP5 GRP6 GRP7

- O N E W A Y -

VARIABLE LENGTH
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	5	5.5815	1.1163	9.376	0
WITHIN GROUPS	395	47.0295	.1191		
TOTAL	400	52.6110			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
GRP 1	68	4.9676	.3747	.0454	3.3000	5.8000	4.8769 TO 5.0584
GRP 2	77	5.2753	.3079	.0351	4.5000	6.1000	5.2054 TO 5.3452
GRP 3	70	4.9743	.3855	.0461	4.1000	5.8000	4.8824 TO 5.0662
GRP 4	73	5.1027	.2749	.0322	4.0000	5.7000	5.0386 TO 5.1669
GRP 5	73	4.9658	.3561	.0417	4.0000	5.6000	4.8827 TO 5.0488
GRP 6	40	5.0000	.3789	.0599	4.0000	5.9000	4.8788 TO 5.1212
TOTAL	401	5.0554			3.3000	6.1000	
	UNGROUPED DATA		.3627	.0181			5.0198 TO 5.0910

Figure IV.1. LeBlanc test A. ANOVA of lengths pooled over beakers within groups.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST A 05/07/81 11.53.49. PAGE 5
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP1

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1.2335	.4112	3.219	.0285
WITHIN GROUPS	64	8.1753	.1277		
TOTAL	67	9.4088			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN		
A	17	4.8000	.4541	.1101	3.9000	5.6000	4.5665	TO	5.0335
B	17	5.1000	.2550	.0618	4.5000	5.6000	4.9689	TO	5.2311
C	17	5.1000	.4016	.0974	4.1000	5.8000	4.8935	TO	5.3065
D	17	4.8706	.2801	.0679	4.4000	5.5000	4.7266	TO	5.0146
TOTAL	68	4.9676			3.9000	5.8000			
	UNGROUPED DATA		.3747	.0454			4.8769	TO	5.0584

Figure IV.2. AVOVA output. LeBlanc test A. Group 1 (Water control).

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST A 05/07/81 11.53.49. PAGE 7
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP2

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.1941	.0647	.674	.5708
WITHIN GROUPS	73	7.0090	.0960		
TOTAL	76	7.2031			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	18	5.3500	.2307	.0544	5.0000	5.7000	5.2353 TO 5.4647
B	20	5.2300	.3114	.0696	4.5000	5.6000	5.0843 TO 5.3757
C	20	5.2300	.3922	.0877	4.5000	6.1000	5.0465 TO 5.4135
D	19	5.3000	.2728	.0626	4.9000	5.7000	5.1685 TO 5.4315
TOTAL	77	5.2753			4.5000	6.1000	
	UNGROUPED DATA		.3079	.0351			5.2054 TO 5.3452

Figure IV.3 ANOVA output. LeBlanc test A. Group 2 (Solvent control).

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1.9310	.6437	5.104	.0031
WITHIN GROUPS	66	8.3227	.1261		
TOTAL	69	10.2537			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	20	5.1850	.2477	.0554	4.6000	5.6000	5.0691 TO 5.3309
B	19	4.8947	.4390	.1007	4.2000	5.7000	4.6831 TO 5.1063
C	15	4.7333	.2823	.0728	4.2000	5.0000	4.5772 TO 4.8895
D	16	5.0313	.4143	.1036	4.1000	5.8000	4.8105 TO 5.2520
TOTAL	70	4.9743			4.1000	5.8000	
	UNGROUPED DATA		.3655	.0461			4.8824 TO 5.0662

Figure IV.4. ANOVA output. LeBlanc test A. Group 3.

TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP4

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.5636	.1879	2.658	.0550
WITHIN GROUPS	69	4.8759	.0707		
TOTAL	72	5.4395			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN		
A	17	5.1647	.2827	.0686	4.6300	5.6000	5.0193	TO	5.3101
B	20	5.0900	.2174	.0486	4.7000	5.5000	4.9883	TO	5.1917
C	17	4.9588	.3501	.0849	4.0000	5.4000	4.7788	TO	5.1388
D	19	5.1895	.2025	.0464	4.5000	5.7000	5.0919	TO	5.2871
TOTAL	73	5.1027			4.0000	5.7000			
	UNGROUPED DATA		.2749	.0322			5.3386	TO	5.1669

Figure IV.5. ANOVA output. LeBlanc test A. Group 4.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST A 05/07/81 11.53.43. PAGE 13
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRPS

- ONEWAY -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.8066	.2689	2.230	.0924
WITHIN GROUPS	69	8.3178	.1205		
TOTAL	72	9.1244			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	18	5.1389	.3712	.0875	4.2000	5.6000	4.3543 TO 5.3235
B	20	4.9600	.3575	.0799	4.0000	5.3000	4.7927 TO 5.1273
C	16	4.8625	.2964	.0741	4.1000	5.3000	4.7046 TO 5.0204
D	19	4.8947	.3519	.0807	4.0000	5.5000	4.7251 TO 5.0644
TOTAL	73	4.9658			4.0000	5.6000	
	UNGROUPED DATA		.3560	.0417			4.8827 TO 5.0488

Figure IV.6. ANOVA output. LeBlanc test A. Group 5.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST A 05/07/81 11.53.43. PAGE 15
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP6

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.1708	.0569	.378	.7697
WITHIN GROUPS	36	5.4292	.1508		
TOTAL	39	5.6000			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	18	4.9889	.4588	.1081	4.3000	5.9000	4.7609 TO 5.2170
B	16	4.9562	.3464	.0866	4.1000	5.5000	4.7717 TO 5.1408
C	5	5.1400	.1140	.0510	5.0000	5.3000	4.9984 TO 5.2816
D	1	5.2000	0	0	5.2000	5.2000	5.2000 TO 5.2000
TOTAL	40	5.0000			4.0000	5.9000	
	UNGROUPED DATA		.3789	.0599			4.8788 TO 5.1212

Figure IV.7. ANOVA output. LeBlanc test A. Group 6.

TREATMENT GROUP AVERAGES
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP1 GRP2 GRP3 GRP4 GRP5 GRP6 GRP7

- O N E W A Y -

VARIABLE LENGTH
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	6	1.7490	.2915	2.013	.0626
WITHIN GROUPS	452	65.4608	.1448		
TOTAL	458	67.2098			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN		
GRP 1	72	4.9528	.3081	.0363	4.0000	5.7000	4.8804	TO	5.0252
GRP 2	67	4.8239	.3722	.0455	4.0000	5.6000	4.7331	TO	4.9147
GRP 3	76	4.8645	.4681	.0537	4.0000	5.8000	4.7575	TO	4.9714
GRP 4	75	4.9360	.3700	.0427	4.1000	5.7000	4.8509	TO	5.0211
GRP 5	67	4.8358	.3171	.0387	4.1000	5.5000	4.7585	TO	4.9132
GRP 6	72	4.8333	.4196	.0494	4.0000	5.8000	4.7347	TO	4.9319
GRP 7	30	4.7267	.3657	.0668	4.0000	5.4000	4.5901	TO	4.8632
TOTAL	459	4.8660			4.0000	5.8000			
	UNGROUPED DATA		.3831	.0179			4.8309	TO	4.9012

Figure IV.8. LeBlanc test B. ANOVA of lengths pooled over beakers within groups.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST B 05/07/81 12.16.57. PAGE 5
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP1

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.1633	.0544	.563	.6413
WITHIN GROUPS	68	6.5761	.1967		
TOTAL	71	6.7394			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	19	5.0253	.3557	.0816	4.2000	5.7000	4.8543 TO 5.1977
B	18	4.9111	.2676	.0631	4.5000	5.5000	4.7780 TO 5.0442
C	20	4.9150	.3483	.0779	4.0000	5.5000	4.7520 TO 5.0780
D	15	4.9600	.2354	.0608	4.5000	5.4000	4.8296 TO 5.0904
TOTAL	72	4.9528			4.0000	5.7000	
	UNGROUPED DATA		.3081	.0363			4.8804 TO 5.0252

Figure IV.9. ANOVA output. LeBlanc test B. Group 1 (Water Control).

TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP2

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1.6418	.5473	4.597	.0057
WITHIN GROUPS	63	7.5000	.1190		
TOTAL	66	9.1418			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN		
A	17	4.8000	.2550	.0618	4.5000	5.2000	4.6689	TO	4.9311
B	16	4.7500	.3483	.0871	4.2000	5.4000	4.5644	TO	4.9356
C	16	4.6500	.3559	.0890	4.0000	5.1000	4.4604	TO	4.8396
D	18	5.0667	.4013	.0946	4.1000	5.6000	4.8670	TO	5.2663
TOTAL	67	4.8239			4.0000	5.6000			
	UNGROUPED DATA		.3722	.0455			4.7331	TO	4.9147

Figure IV.10. ANOVA output. LeBlanc test B. Group 2 (Solvent control).

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST B 05/07/81 12.15.57. PAGE 9
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP3

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1.6130	.5377	2.612	.0578
WITHIN GROUPS	72	14.8211	.2058		
TOTAL	75	16.4341			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	20	4.7900	.5015	.1121	4.0000	5.5000	4.5553 TO 5.0247
B	18	4.9944	.4080	.0962	4.5000	5.8000	4.7915 TO 5.1973
C	20	4.6750	.4153	.0929	4.0000	5.7000	4.4806 TO 4.8694
D	18	5.0278	.4812	.1134	4.1000	5.8000	4.7885 TO 5.2671
TOTAL	76	4.8645			4.0000	5.8000	
	UNGROUPED DATA		.4681	.0537			4.7575 TO 4.9714

Figure IV.11. ANOVA output. LeBlanc test B. Group 3.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST B 05/07/81 12.16.57. PAGE 11
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP4

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.3889	.1296	.945	.4239
WITHIN GROUPS	71	9.7439	.1372		
TOTAL	74	10.1328			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	17	4.9647	.3427	.0831	4.3000	5.5000	4.7885 TO 5.1409
B	20	5.0200	.3270	.0731	4.5000	5.7000	4.8669 TO 5.1731
C	20	4.9300	.3881	.0868	4.1000	5.7000	4.7484 TO 5.1116
D	18	4.8222	.4181	.0985	4.2000	5.5000	4.6143 TO 5.0301
TOTAL	75	4.9360			4.1000	5.7000	
	UNGROUPED DATA		.3700	.0427			4.8509 TO 5.0211

Figure IV.12. ANOVA output. LeBlanc test B. Group 4.

DESCRIPTIVE STATS FOR DAPHNIA LENGTHS, BIONOMICS TEST B 05/07/81 12.16.57. PAGE 13
TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRPS

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.1878	.0626	.612	.6099
WITHIN GROUPS	63	6.4462	.1023		
TOTAL	66	6.6340			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	18	4.8389	.3648	.0860	4.2000	5.3000	4.6575 TO 5.0203
B	17	4.8059	.2633	.0639	4.2000	5.0000	4.6705 TO 4.9413
C	15	4.7733	.2631	.0679	4.4000	5.3000	4.6276 TO 4.9191
D	17	4.9176	.3627	.0880	4.1000	5.5000	4.7312 TO 5.1041
TOTAL	67	4.8358			4.1000	5.5000	
	UNGROUPED DATA		.3170	.0387			4.7585 TO 4.9132

Figure IV.13. ANOVA output. LeBlanc test B. Group 5.

TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP6

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1.4236	.4745	2.913	.0405
WITHIN GROUPS	68	11.0764	.1629		
TOTAL	71	12.5000			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	18	4.9333	.4839	.1140	4.1000	5.8000	4.6927 TO 5.1739
B	20	4.6300	.3279	.0733	4.0000	5.3000	4.4765 TO 4.7834
C	16	4.8062	.3672	.0968	4.1000	5.5000	4.5999 TO 5.0126
D	18	4.9633	.4062	.0957	4.2000	5.4000	4.7813 TO 5.1453
TOTAL	72	4.8333			4.0000	5.8000	
	UNGROUPED DATA		.4196	.0494			4.7347 TO 4.9319

Figure IV.14. ANOVA output. LeBlanc test B. Group 6.

TANK AVERAGES FOR EACH TREATMENT GROUP (SUBFILE)
FILE NONAME (CREATION DATE = 05/07/81)
SUBFILE GRP7

- O N E W A Y -

VARIABLE LENGTH
BY TANK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	.2657	.0886	.637	.5978
WITHIN GROUPS	26	3.6130	.1390		
TOTAL	29	3.8787			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
A	3	4.5667	.4163	.2404	4.1000	4.9000	3.5324 TO 5.6009
B	14	4.7143	.3159	.0844	4.1000	5.3000	4.5319 TO 4.6967
C	12	4.8083	.4231	.1221	4.0000	5.4000	4.5395 TO 5.0772
D	1	4.4000	0	0	4.4000	4.4000	4.4000 TO 4.4000
TOTAL	30	4.7267			4.0000	5.4000	
	UNGROUPED DATA		.3657	.0668			4.5901 TO 4.8632

Figure IV.15. ANOVA output. LeBlanc test B. Group 7.

V. ADJUSTMENTS TO ACCOUNT FOR BEAKER TO BEAKER HETEROGENEITY WITHIN TREATMENT GROUPS

A. BACKGROUND

In the previous two sections we have discussed procedures to test for beaker to beaker heterogeneity within treatment groups for survival and for length responses. In both sections we found statistical evidence of beaker to beaker heterogeneity. In this section we discuss either adjustments in the data or adjustments in the statistical procedures used, in order to account for this heterogeneity. As remarked in the previous section, we need not adjust the productivity data since this is measured and reported on a per beaker basis and there is no obvious theoretical model upon which to base variability estimates. We thus carry out statistical analyses on a per beaker basis. In the subsections below we discuss procedure to adjust the mortality responses and the length responses prior to comparison across groups.

B. ADJUSTMENTS TO ACCOUNT FOR HETEROGENEITY OF MORTALITY RESPONSES

We tested for beaker to beaker heterogeneity in mortality in Section III. We found some evidence of heterogeneity in the data from LeBlanc's Tests A and B. There was no heterogeneity of responses in Adams' data (also virtually no partial kills). Chapman's data consists of just a single daphnid per beaker, so beaker to beaker variation cannot be determined. Goulden's isophorone data show no evidence of heterogeneity of responses within groups, among the beakers with multiple daphnids. We thus need to account for beaker to beaker heterogeneity in mortality in LeBlanc's data but not in Adams', Chapman's or Goulden's data.

The most commonly used approach for the comparison of mortality rates across treatment groups is to carry out an arc sin variance stabilizing transformation on the observed mortality rate within each beaker and then use the mean square for variation among beakers within groups (pooled over all groups) as an error yardstick. If there are I groups, J beakers per group, then this error yardstick has I(J-1) degrees of freedom associated with it.

This is a conservative approach. Although the beakers within groups mean square is correct whether or not beaker to beaker variation in mortality exists, it is based on relatively few degrees of freedom and thus can lead to diminished sensitivity of inferences (i.e. lowered power of tests, increased lengths of confidence intervals) if I(J-1) is small. For example if I = 6, J = 3 then I(J-1) = 12. The sensitivities of the analyses would be improved if the degrees of freedom for the error yardstick could somehow be increased.

Feder and Collins [1], Section IX discuss an approach to accounting for the increased variability introduced by tank to tank heterogeneity in fathead minnow mortality rates by reducing actual sample sizes per group to effective sample sizes and than disregarding the tank effects.

A very similar situation exists for the Daphnia mortality responses and so we adopt a similar approach here. Following Feder and Collins, suppose that there are 20 daphnids per beaker and a "true" mortality rate of p. Under binomial theory the variance of $\hat{p}$, the estimated mortality rate, should be p(1-p)/20. Suppose however that the responses within each beaker are positively correlated due to beaker to beaker heterogeneity and this increases the variance of $\hat{p}$ by 20 percent to 1.2p(1-p)20. Then we can regard the effective sample size within that beaker as 20/1.2=16.67. To maintain the observed response rate at its level p we adjust both the number dead and the number live by the same factor. For example if the data as reported show 5 deaths in 20 daphnids, we would ajdust this down to 5/1.2 = 4.17 deaths in 20/1.2 = 16.67 daphnids. Beaker to beaker heterogeneity is then ignored, effective numbers of responses and daphnids are pooled across beakers within groups, and standard binomial based procedures are applied to the adjusted data as if no beaker to beaker variation within groups existed.

We now consider the calculation of adjustment factors. Adjustment factors can be calculated separately for each group or a single adjustment factor can be calculated for all the groups combined. We first consider the calculation of a single adjustment factor and then consider displays that suggest whether single or separate adjustments are called for.

Motivation for the adjustment procedure comes from the form of the beta binomial model [Williams [3]]. Suppose X_{ij} is the number of responses in beaker j of group i (e.g. number dead after 21 days). The beta binomial model extends the binomial to allow for beaker to beaker variation within groups. Following Feder and Collins [1], Section IX we assume that $X_{ij} \sim$ Binomial (N_{ij}, p_{ij}) where N_{ij} is the sample size within beaker j of group i and p_{ij} is the response probability there. Further, it is assumed that $p_{ij} \sim$ Beta (α_i, β_i) where α_i and β_i are unknown parameters. Let

$$\mu_i = \frac{\alpha_i}{\alpha_i + \beta_i} \qquad \theta_i = \frac{1}{\alpha_i + \beta_i}$$

Then it is known that X_{ij} has a beta binomial distribution with parameters $(N_{ij}, \mu_i, \theta_i)$. See Williams for details. In particular it can be shown directly that

$$E(X_{ij}) = N_{ij}\,\mu_i$$

$$\mathrm{Var}(X_{ij}) = N_{ij}\,\mu_i(1-\mu_i)\left[\frac{1 + N_{ij}\,\theta_i}{1 + \theta_i}\right] \qquad 0 \leq \theta_i < \infty$$

We see that the variance of X_{ij} is inflated over and above binomial variance by a multiplicative factor.

Suppose that $N_{ij} \equiv N_i$, $j=1,\ldots, J_i$. This assumption is often reasonable in Daphnia tests, where N_{ij} represents the number of daphnids in the beaker at the outset of the test. In fact the assumption $N_{ij} \equiv N$ is often reasonable also. (Note that in tests containing beakers with single daphnids and beakers with multiple daphnids, we are referring just to the beakers with multiple daphnids. e.g. In Goulden's test, this discussion would refer just the three beakers per group containing five daphnids each.) Then the multiplicative factor is $(1 + N_i \theta_i)/(1 + \theta_i) \equiv K_i$, $j=1,\ldots, J_i$. Thus $\mathrm{Var}(X_{ij}) = N_i K_i \mu_i(1-\mu_i)$ where $1 \leq K_i < \infty$. Define $\hat{p}_{ij} \equiv X_{ij}/N_i$. $\hat{p}_{ij}$ is the observed response proportion. Therefore

$$\mathrm{Var}(\hat{p}_{ij}) = \frac{K_i}{N_i} \mu_i(1-\mu_i) \qquad j=1,\ldots, J_i$$

Thus the effective sample size per beaker in the i-th group is N_i/K_i. As the extent of beaker to beaker variation approaches 0 (i.e. as $\theta_i \to 0$), K_i approaches 1 and so N_i/K_i approaches N_i. As the extent of beaker to beaker variation gets greater and greater (i.e. as $\theta_i \to \infty$), K_i approaches N_i and so N_i/K_i approaches 1. These two extreme situations call for carrying out analyses on a per daphnid basis (after pooling responses across beakers within groups) or carrying out analyses on a per beaker basis. In general some middle ground is appropriate. Note that if $N_i \equiv N$ and $\theta_i \equiv \theta$ for all i, then $K_i \equiv K$ for all i and K should be estimated based on results from all the groups.

The procedure discussed below for calculating adjustment factors is motivated by the beta binomial theory results. A full fledged maximum likelihood procedure to fit the beta binomial model to the data might be developed, but we decided to use a simpler procedure based on the method of moments.

Let X_{ij}, N_i denote the number of responses and the total number of daphnids respectively within beaker j of group i. Let $\hat{p}_{ij} \equiv X_{ij}/N_i$. The variance inflation factor, K_i, is defined as

$$K_i \equiv \frac{\mathrm{Var}(\hat{p}_{ij})}{[\mu_i(1-\mu_i)/N_i]}$$

This suggests that K_i be estimated by substituting the sample analogues of $\mathrm{Var}(\hat{p}_i)$ and μ_i in the expression above. If the N_{ij}'s are not in fact exactly the same within each beaker then use the average sample size in the expression for K_i. (Use the harmonic average if the N_{ij}'s differ by much.)

Suppose there are I treatment and control groups, J_i beakers per group. Within the i-th group let $N_i = J_i^{-1} \Sigma_j N_{ij}$, $\hat{p}_i \equiv J_i^{-1} \Sigma_j \hat{p}_{ij}$, $\hat{\mathrm{Var}}(\hat{p}_{ij}) \equiv (J_i-1)^{-1} \Sigma_j(\hat{p}_{ij} - \hat{\bar{p}}_i)^2$ denote the average sample size, the average observed

response rate, and the sample variance of response rates respectively. The N_{ij}'s are generally nearly equal, if not exactly equal, in daphnid toxicity test data. We estimate K_i as

$$\hat{K}_i = \hat{V}ar(\hat{p}_{ij})/[\bar{\hat{p}}_i(1-\bar{\hat{p}}_i)/\bar{N}_i]$$

We either pool the $\hat{K}_i$'s across groups to obtain an overall adjustment factor, $\hat{K}$, or else we use separate adjustment factors within each group. We will describe later in this subsection a graphical procedure for comparing the K_i's across groups. For now we consider an overall adjustment factor based on all the groups.

Let $\hat{K}_i$ denote the inflation factor within the i-th group. If we take $\bar{\hat{p}}_i$ as essentially μ_i (this is reasonable, unless μ_i is very close to 0 or to 1, since $\bar{\hat{p}}_i$ is an average value over all the daphnids in the group) then $\hat{K}_i$ can be regarded as distributed approximately as $K_i\ \chi^2_{(J_i-1)}/(J_i-1)$. If $K_i \equiv K$, i=1,..., I then we can obtain a pooled estimate of K as

$$\bar{\hat{K}} = \Sigma_i(J_i-1)\ \hat{K}_i/\Sigma_i(J_i-1)$$

If $J_i \equiv J$ for all i then the expression for $\hat{K}$ reduces to the simple average, $\Sigma_i\ K_i/I$. Under the above assumptions, $\bar{\hat{K}}$ is approximately distributed as $K\ \chi^2_{\Sigma_i(J_i-1)}/\Sigma_i(J_i-1)$. This distributional result can be used to provide an upper confidence bound on K. Namely, a 95 percent upper confidence bound on K is

$$K \leq \bar{\hat{K}}\ \Sigma_i(J_i-1)/\chi^{2\ (0.05)}_{\Sigma_i(J_i-1)}$$

where $\chi^{2\ (0.05)}_{\Sigma_i(J_i-1)}$ is the 5th percentile of the chi square distribution with $\Sigma_i(J_i-1)$ d.f. Denote this upper confidence bound by $\tilde{K}_u$.

The suggested adjustment procedure is to reduce the effective sample size within the ij-th beaker to $N_{ij}/\hat{K}$ in such a way that $\hat{p}_{ij}$ is unchanged. (The ratio $N_{ij}/\hat{K}$ is constrained to lie between 1 and N_{ij}.) Then ignore beaker to beaker variation, pool results across beakers within groups, and carry out subsequent analyses based on pooled results within groups.

When comparing treatment effects based on the adjusted data, how many degrees of freedom should be associated with the error yardstick corresponding to variation among daphnids within groups? If we use binomial theory, then we are assuming an infinite number of degrees of freedom for this yardstick. This would be appropriate if and only if K were known, which is not usually the case. At the other extreme, we could argue that $\bar{\hat{K}}$ is based on $\Sigma_i(J_i-1)$ degrees of freedom and so this should be associated with the error yardstick. If $J_i = J$ for all i, this becomes I(J-1) degrees of freedom. Adjusting the sample sizes by $\hat{K}$ and then using $\Sigma_i(J_i-1)$ or I(J-1) degrees of freedom as the case may be, is very close to Finney's [4] suggestion of using a heterogeneity factor based

on the residual mean square from the (probit) model fit. This is a conservative viewpoint and uses no information about the relation between the observed beaker to beaker variation and that predicted by binomial theory. Diminished sensitivity can result when I(J-1) is small. This can happen particularly if J=2 or 3.

A middle ground between the two extremes discussed above could be based on reasoning as follows. (Assume for the purpose of discussion that $N_{ij} \equiv N$, $J_i \equiv J$. This simplifies notation. However the ideas are more general.) Since we are assuming an effective sample size of JN/K per group and pooling data across beakers, each group provides JN/K-1 degrees of freedom for estimating variability. Now $\tilde{K}_u$ is an upper confidence bound on K. Since K is unknown, we substitute this upper bound for it and thus assume $JN/\tilde{K}_u-1$ degrees of freedom per group, or $I(JN/\tilde{K}_u-1)$ degrees of freedom altogether. If $\tilde{K}_u = 1$ then we have I(JN-1) degrees of freedom and we have made no adjustment. If $\tilde{K}_u = N$ then there is effectively one observation per beaker and so we have I(J-1) degrees of freedom, just as we associated with per beaker analyses or with using a heterogeneity factor. The ratio $I(JN/\tilde{K}_u-1)$ is constrained to lie between I(J-1) and I(JN-1).

As an alternative to adjusting the sample sizes to effective sample sizes, $N/\hat{K}$, we can carry out comparisons on a per beaker basis (after performing an arc sin variance stablizing transformation) and use the mean square for beakers within groups as an error yardstick. The usual practice is to associate I(J-1) degrees of freedom with this yardstick. A less conservative practice would be to compare the magnitude of this yardstick with that expected based on binomial theory and pool this information to arrive at the increased number of degrees of freedom $I(JN/\tilde{K}_u-1)$. This will increase the sensitivity of comparisons among treatment effects, particularly when I(J-1) is rather small. Note that $N/\tilde{K}_u$ is constrained to lie between 1 and N.

We now apply this procedure to the mortality data from LeBlanc's Tests A and B. From the preliminary tests of beaker to beaker heterogeneity in Section III we concluded that there is strong statistical evidence of heterogeneity in each test. (However the heterogeneity may be in just a single group in each case.)

LeBlanc Test A

Group 1: $\hat{p}_{11} = 0.15$, $\hat{p}_{12} = 0.15$, $\hat{p}_{13} = 0.15$, $\hat{p}_{14} = 0.15$, $\hat{\bar{p}}_1 = 0.15$,

$N_{11} = N_{12} = N_{13} = N_{14} = \bar{N}_1 = 20$, $\hat{\text{Var}}(\hat{p}_{ij}) = 0.00$,

$\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1 = 0.00638$

$$\hat{K}_1 = \frac{\hat{Var}(\hat{p}_{ij})}{[\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1]} = \frac{0.00}{0.00638} = 0.00$$

Group 2: $\hat{p}_{21} = 0.10$, $\hat{p}_{22} = 0.00$, $\hat{p}_{23} = 0.00$, $\hat{p}_{24} = 0.05$, $\hat{\bar{p}}_2 = 0.0375$,
$N_{21} = N_{22} = N_{23} = N_{24} = \bar{N}_2 = 20$, $\hat{Var}(\hat{p}_{2j}) = 0.0023$,
$\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2 = 0.00180$

$$\hat{K}_2 = \frac{\hat{Var}(\hat{p}_{2j})}{[\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2]} = \frac{0.0023}{0.0018} = 1.28$$

Group 3: $\hat{p}_{31} = 0.00$, $\hat{p}_{32} = 0.05$, $\hat{p}_{33} = 0.25$, $\hat{p}_{34} = 0.20$, $\hat{\bar{p}}_3 = 0.125$,
$N_{31} = N_{32} = N_{33} = N_{34} = \bar{N}_3 = 20$, $\hat{Var}(\hat{p}_{3j}) = 0.0142$,
$\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3 = 0.0055$

$$\hat{K}_3 = \frac{\hat{Var}(\hat{p}_{3j})}{[\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3]} = \frac{0.0142}{0.0055} = 2.58$$

Group 4: $\hat{p}_{41} = 0.15$, $\hat{p}_{42} = 0.00$, $\hat{p}_{43} = 0.15$, $\hat{p}_{44} = 0.05$, $\hat{\bar{p}}_4 = 0.0875$,
$N_{41} = N_{42} = N_{43} = N_{44} = \bar{N}_4 = 20$, $\hat{Var}(\hat{p}_{4j}) = 0.0056$,
$\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/N_4 = 0.0040$

$$\hat{K}_4 = \frac{\hat{Var}(\hat{p}_{4j})}{[\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4]} = \frac{0.0056}{0.0040} = 1.40$$

Group 5: $\hat{p}_{51} = 0.10$, $\hat{p}_{52} = 0.00$, $\hat{p}_{53} = 0.20$, $\hat{p}_{54} = 0.05$, $\hat{\bar{p}}_5 = 0.0875$,
$N_{51} = N_{52} = N_{53} = N_{54} = \bar{N}_5 = 20$, $\hat{Var}(\hat{p}_{5j}) = 0.0073$,
$\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5 = 0.0040$

$$\hat{K}_5 = \frac{\hat{Var}(\hat{p}_{5j})}{[\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5]} = \frac{0.0073}{0.0040} = 1.825$$

Group 6: $\hat{p}_{61} = 0.10$, $\hat{p}_{62} = 0.20$, $\hat{p}_{63} = 0.75$, $\hat{p}_{64} = 0.95$, $\hat{\bar{p}}_6 = 0.50$,

$N_{61} = N_{62} = N_{63} = N_{64} = \bar{N}_6 = 20$, $\hat{Var}(\hat{p}_{6j}) = 0.1717$,

$\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6 = 0.0125$

$$\hat{K}_6 = \frac{\hat{Var}(\hat{p}_{6j})}{[\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6]} = \frac{0.1717}{0.0125} = 13.76$$

Group 7: $\hat{p}_{71} = 1.00$, $\hat{p}_{72} = 1.00$, $\hat{p}_{73} = 0.95$, $\hat{p}_{74} = 1.00$, $\hat{\bar{p}}_7 = 0.9875$,

$N_{71} = N_{72} = N_{73} = N_{74} = \bar{N}_7 = 20$, $\hat{Var}(\hat{p}_{7j}) = 0.00063$,

$\hat{\bar{p}}_7(1-\hat{\bar{p}}_7)/\bar{N}_7 = 0.00062$

$$\hat{K}_7 = \frac{\hat{Var}(\hat{p}_{7j})}{[\hat{\bar{p}}_7(1-\hat{\bar{p}}_7)/\bar{N}_7]} = \frac{0.00063}{0.00062} = 1.02$$

It is obvious that the inflation factor from group 6 dominates all the others. However if we ignore this for the moment for the sake of illustration and calculate an overall inflation factor, we obtain

$$\hat{\bar{K}} = \frac{\Sigma_{i=1}^{7} \hat{K}_i}{7} = \frac{0.00 + 1.28 + 2.58 + 1.40 + 1.83 + 13.76 + 1.02}{7} = 3.12$$

Under the (rather dubious) assumption that all the K_i's are equal, the distribution of K may be approximated as $K\,\chi^2_{I(J-1)}/I(J-1) = K\,\chi^2_{21}/21$. A 95 percent upper confidence bound on K would then be

$$K \leq \hat{\bar{K}}\, I(J-1)/\chi^2_{I(J-1)}(0.05) = (3.12)(21)/\chi^2_{21}(0.05) = (3.12)(21)/11.6 = 5.65$$

The suggested adjustment procedure is to reduce the effective sample size within each beaker to 20/3.12 = 6.41 while maintaining the observed mortality rates, disregard beaker to beaker variation and pool results across beakers within groups, and carry out subsequent analyses based on the pooled results within groups.

When comparing treatment effects we associate $I(JN/\tilde{K}_u-1) = 7((4)(20)/5.65-1) = 92.1 \simeq 92$ degrees of freedom with the error yardstick.

LeBlanc Test B

Group 1: $\hat{p}_{11} = 0.05$, $\hat{p}_{12} = 0.10$, $\hat{p}_{13} = 0.00$, $\hat{p}_{14} = 0.25$, $\hat{\bar{p}}_1 = 0.10$,

$N_{11} = N_{12} = N_{13} = N_{14} = \bar{N}_1 = 20$, $\hat{Var}(\hat{p}_{1j}) = 0.0117$,

$\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1 = 0.0045$

$$\hat{K}_1 = \frac{\hat{Var}\ (\hat{p}_{1j})}{[\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1]} = \frac{0.0117}{0.0045} = 2.60$$

Group 2: $\hat{p}_{21} = 0.15$, $\hat{p}_{22} = 0.20$, $\hat{p}_{23} = 0.20$, $\hat{p}_{24} = 0.10$, $\hat{\bar{p}}_2 = 0.1625$,

$N_{21} = N_{22} = N_{23} = N_{24} = \bar{N}_2 = 20$, $\hat{Var}(\hat{p}_{2j}) = 0.0023$,

$\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2 = 0.0068$

$$\hat{K}_2 = \frac{\hat{Var}(\hat{p}_{2j})}{[\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2]} = \frac{0.0023}{0.0068} = 0.3368$$

Group 3: $\hat{p}_{31} = 0.00$, $\hat{p}_{32} = 0.10$, $\hat{p}_{33} = 0.00$, $\hat{p}_{34} = 0.10$, $\hat{\bar{p}}_3 = 0.05$,

$N_{31} = N_{32} = N_{33} = N_{34} = \bar{N}_3 = 20$, $\hat{Var}(\hat{p}_{3j}) = 0.0033$,

$\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3 = 0.0024$

$$\hat{K}_3 = \frac{\hat{Var}(\hat{p}_{3j})}{[\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3]} = 1.375$$

Group 4: $\hat{p}_{41} = 0.15$, $\hat{p}_{42} = 0.00$, $\hat{p}_{43} = 0.00$, $\hat{p}_{44} = 0.10$, $\hat{\bar{p}}_4 = 0.0625$,

$N_{41} = N_{42} = N_{43} = N_{44} = \bar{N}_4 = 20$, $\hat{Var}(\hat{p}_{4j}) = 0.0056$,

$\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4 = 0.0029$

$$\hat{K}_4 = \frac{\hat{Var}(\hat{p}_{4j})}{[\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4]} = \frac{0.0056}{0.0029} = 1.93$$

Group 5: $\hat{p}_{51} = 0.10$, $\hat{p}_{52} = 0.15$, $\hat{p}_{53} = 0.25$, $\hat{p}_{54} = 0.15$, $\hat{\bar{p}}_5 = 0.1625$,

$N_{51} = N_{52} = N_{53} = N_{54} = \bar{N}_5 = 20$, $\hat{\text{Var}}(\hat{p}_{5j}) = 0.0040$,

$\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5 = 0.0068$

$$\hat{K}_5 = \frac{\hat{\text{Var}}(\hat{p}_{5j})}{[\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5]} = \frac{0.0040}{0.0068} = 0.588$$

Group 6: $\hat{p}_{61} = 0.10$, $\hat{p}_{62} = 0.00$, $\hat{p}_{63} = 0.20$, $\hat{p}_{64} = 0.10$, $\hat{\bar{p}}_6 = 0.10$,

$N_{61} = N_{62} = N_{63} = N_{64} = \bar{N}_6 = 20$, $\hat{\text{Var}}(\hat{p}_{6j}) = 0.0067$,

$\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6 = 0.0045$

$$\hat{K}_6 = \frac{\hat{\text{Var}}(\hat{p}_{6j})}{[\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6]} = \frac{0.0067}{0.0045} = 1.49$$

Group 7: $\hat{p}_{71} = 0.85$, $\hat{p}_{72} = 0.30$, $\hat{p}_{73} = 0.40$, $\hat{p}_{74} = 0.95$, $\hat{\bar{p}}_7 = 0.625$,

$N_{71} = N_{72} = N_{73} = N_{74} = \bar{N}_7 = 20$, $\hat{\text{Var}}(\hat{p}_{7j}) = 0.1042$,

$\hat{\bar{p}}_7(1-\hat{\bar{p}}_7)/\bar{N}_7 = 0.0117$

$$\hat{K}_7 = \frac{\hat{\text{Var}}(\hat{p}_{7j})}{[\hat{\bar{p}}_7(1-\hat{\bar{p}}_7)/\bar{N}_7]} = \frac{0.1042}{0.0117} = 8.906$$

It is obvious that the inflation factor from group 7 dominates all the others. However, we ignore this for the moment for the sake of illustration and calculate an overall inflation factor.

$$\hat{\bar{K}} = \frac{\sum_{i=1}^{7} K_i}{7} = \frac{2.60 + 0.34 + 1.384 + 1.93 + 0.59 + 1.49 + 8.91}{7} = 2.46$$

Under the (dubious) assumption that all the K_i's are equal, the distribution of $\hat{\bar{K}}$ may be approximated as $K\,\chi^2_{I(J-1)}/I(J-1) = K\,\chi^2_{21}/21$. A 95 percent upper confidence bound on K would then be

$$K \leq \hat{\bar{K}}\, I(J-1)/\chi^2_{I(J-1)}(0.05) = (2.46)(21)/11.6 = 4.45$$

The suggested adjustment procedure is to reduce the effective sample size within each beaker to 20/2.46 = 8.13 while maintaining the observed

mortality rates, disregard beaker to beaker variation and pool results across beakers within groups, and carry out subsequent analyses based on the pooled results within groups.

When comparing treatment effects we associate $I(JN/\tilde{K}_u-1) = 7((4)(20)/4.45-1) = 118.8 \approx 119$ degrees of freedom with the error yardstick.

The model upon which adjustment calculations were based assumes that the extent of extrabinomial variation (i.e. the K_i's) is constant across groups. Before applying adjustments based on this assumption one should determine whether it is realistic. There were indications, mentioned above, that the significant beaker to beaker heterogeneity observed in LeBlanc's Tests A and B may all be due to the extreme heterogeneity in group 6 for Test A or group 7 for Test B respectively. If the heterogeneity among beakers differs from group to group, then separate adjustment factors should be used within each group.

We discuss below a procedure for determining whether the extent of observed beaker to beaker extrabinomial variation is the same for all groups or is greater for some groups than for others. The procedure consists of calculating separate inflation factors, $\hat{K}_i$, within each group and comparing them across groups. Under the assumption of a common theoretical inflation factor K across groups, these estimated inflation factors are distributed approximately as $K\,\chi^2_{(J-1)}/(J-1)$ (unless p_i is very close to 0 or to 1). We should thus see a straight line or at least a smooth curve when the ordered inflation factors are plotted on chi square probability paper. If one or two points are far removed from the others, this suggests differing amounts of extrabinomial variation across groups. This situation would need to be reflected in subsequent analyses.

The results for LeBlanc's Tests A and B are summarized below in Tables V.1 and V.2. In these tests J=4 and I=7. The rank is the order of the inflation factor, from smallest to largest. The plotting position for the inflation factor with rank i is 100(i-0.5)/7.

TABLE V.1. LE BLANC TEST A--VARIANCE INFLATION FACTOR BY GROUP

Group	Factor ($\hat{K}_i$)	Rank	Plotting Position
1	0.00	1	7.14
2	1.28	3	35.71
3	2.58	6	78.57
4	1.40	4	50.00
5	1.83	5	64.29
6	13.76	7	92.86
7	1.02	2	21.43

TABLE V.2. LE BLANC TEST B--VARIANCE INFLATION FACTOR BY GROUP

Group	Factor (K_i)	Rank	Plotting Position
1	2.60	6	78.57
2	0.34	1	7.14
3	1.38	3	35.71
4	1.93	5	64.29
5	0.59	2	21.43
6	1.49	4	50.00
7	8.91	7	92.86

The inflation factors are plotted versus their plotting positions on chi square probability paper with 3 d.f. The results for Tests A and B are shown in Figures V.1 and V.2 respectively. The reference lines in those plots correspond to the theoretical c.d.f. with distribution $1/3\ \chi^2_3$. The two plots look remarkably similar. The factors corresponding to group 6 in Test A and group 7 in Test B are substantially out of line with those from the other groups. The probability that the maximum of 7 independent random variables, each distributed as $1/3\ \chi^2_3$, exceeds 8.91 is 5×10^{-5} and the probability that it exceeds 13.76 is 4×10^{-8}. Thus these extreme factors are certainly incompatible with beaker to beaker homogeneity and appear to be incompatible with the heterogeneity observed in the remaining groups. Apart from the extreme factors, the remaining groups appear to exhibit beaker to beaker heterogeneity in excess of that to be expected on the basis of binomial theory. The factors all seem comparable across groups. The average inflation factors, excluding the extreme groups are 1.35 for Test A and 1.39 for Test B. The test statistic, Z, in EXAX2 that tests the hypothesis of overall beaker to beaker heterogeneity is 1.00 in Test A and 1.165 in Test B after the extreme groups have been separated. Under the hypothesis of no beaker to beaker heterogeneity, Z has a standard normal distribution. Thus Z is significant at the 16 percent level in Test A and at the 12 percent level in Test B. These results, in agreement with those in Figures V.1 and V.2 are suggestive of some beaker to beaker heterogeneity but do not provide strong statistical evidence. Two reasonable approaches would be to adjust the data in the extreme groups by an inflation factor based only on the responses from those groups and then either apply no adjustment to the remaining groups or else apply the adjustment factors calculated immediately above to these groups. The second approach is slightly more conservative than the first and we adopt it here.

The suggested adjustment procedures for Tests A and B are as follows:

Test A

Group 6: $\hat{K}_6 = 13.76$, $\tilde{K}_{6,u} = \hat{K}_6(J-1)/\chi^2_{(J-1)}(0.05) = (13.76)(3)/\chi^2_3(0.05)$

$= (13.76)(3)/0.352 = 117.27$

Remaining groups: $\hat{\bar{K}} = 1.35$, $\tilde{K}_u = \hat{\bar{K}}(I-1)(J-1)/\chi^2_{(I-1)(J-1)}(0.05)$

$= (1.35)(18)/9.39 = 2.59$

Thus we reduce the effective sample size per beaker in group 6 to 20/13.76 = 1.45 while maintaining the observed mortality rates. We reduce the effective sample size per beaker in the other groups to 20/1.35 = 14.8 while maintaining the observed mortality rates. The degrees of freedom per group are max $[(JN/\tilde{K}_u-1), (J-1)]$. We thus associate 3 degrees of freedom with group 6 and $(4)(20)/2.59-1 = 29.9 \approx 30$ degrees of freedom with each of the other groups.

Test B

Group 7: $\hat{K}_7 = 8.91$, $\tilde{K}_{7,u} = \hat{K}_7(J-1)/\chi^2_{(J-1)}(0.05) = (8.91)(3)/0.352 = 75.94$

Remaining groups: $\hat{\bar{K}} = 1.39$, $\tilde{K}_u = \hat{\bar{K}}(I-1)(J-1)/\chi^2_{(I-1)(J-1)}(0.05)$

$= (1.39)(18)/9.39 = 2.66$

Thus we reduce the effective size per beaker in group 7 to 20/8.91 = 2.24 while maintaining the observed mortality rates. We reduce the effective sample size per beaker in the other groups to 20/1.39 = 14.4 while maintaining the observed mortality rates. We associate 3 degrees of freedom with group 7 and $(4)(20)/2.66 - 1 = 29.1 \approx 29$ degrees of freedom with each of the other groups.

The results of the adjustment procedures applied to the data from Tests A and B are presented in Tables V.3 and V.4 respectively. These adjusted values are used as basic input "data" for subsequent analyses. We then proceed as if there is no beaker to beaker heterogeneity within groups. The extrabinomial variation has been accounted for by the adjustment procedure.

C. ALTERNATIVE MEASURES OF MORTALITY

In the previous subsection we calculated adjustment factors to account for beaker to beaker heterogeneity in mortality responses within groups. All the examples considered there pertained to 21 day mortality. However in Section III we noted that there are alternative measures of mortality which are of importance to study such as 7 day mortality, 21 day mortality conditional on 7 day survival, etc. Each such measure provides information about mortality during different life stages and thus helps to distinguish among the various causes of mortality--biological and experimental.

It is not clear a priori whether or not the extent of beaker to beaker heterogeneity within groups observed and adjusted for in the 21 day mortality responses is also applicable for other mortality responses. If that is the case then the adjustment factors calculated for 21 day mortality

TABLE V.3. EFFECTIVE SAMPLE SIZES AND RESPONSES IN LE BLANC TEST A 21 DAY MORTALITY DATA AFTER ADJUSTMENT FOR BEAKER TO BEAKER HETEROGENEITY

Group		Beaker A	Beaker B	Beaker C	Beaker D	
1	Dead	2.2	2.2	2.2	2.2	
	Live	12.6	12.6	12.6	12.6	
	Total	14.8	14.8	14.8	14.8	30 d.f.
2	Dead	1.5	0.0	0.0	0.7	
	Live	13.3	14.8	14.8	14.1	
	Total	14.8	14.8	14.8	14.8	30 d.f.
3	Dead	0.0	0.7	3.7	3.0	
	Live	14.8	14.1	11.1	11.8	
	Total	14.8	14.8	14.8	14.8	30 d.f.
4	Dead	2.2	0.0	2.2	0.7	
	Live	12.6	14.8	12.6	14.1	
	Total	14.8	14.8	14.8	14.8	30 d.f.
5	Dead	1.5	0.0	3.0	0.7	
	Live	13.3	14.8	11.8	14.1	
	Total	14.8	14.8	14.8	14.8	30 d.f.
6	Dead	0.14	0.29	1.09	1.38	
	Live	1.31	1.16	0.36	0.07	
	Total	1.45	1.45	1.45	1.45	3 d.f.
7	Dead	14.8	14.8	14.1	14.8	
	Live	0.0	0.0	0.7	0.0	
	Total	14.8	14.8	14.8	14.8	30 d.f.

TABLE V.4. EFFECTIVE SAMPLE SIZES AND RESPONSES IN LE BLANC TEST B 21 DAY MORTALITY DATA AFTER ADJUSTMENT FOR BEAKER TO BEAKER HETEROGENEITY

Group		Beaker A	Beaker B	Beaker C	Beaker D	
1	Dead	0.7	1.4	0.0	3.6	
	Live	13.7	12.9	14.4	10.8	
	Total	14.4	14.4	14.4	14.4	29 d.f.
2	Dead	2.2	2.9	2.9	1.4	
	Live	12.2	11.5	11.5	12.9	
	Total	14.4	14.4	14.4	14.4	29 d.f.
3	Dead	0.0	1.4	0.0	1.4	
	Live	14.4	12.9	14.4	12.9	
	Total	14.4	14.4	14.4	14.4	29 d.f.
4	Dead	2.2	0.0	0.0	1.4	
	Live	12.2	14.4	14.4	12.9	
	Total	14.4	14.4	14.4	14.4	29 d.f.
5	Dead	1.4	2.2	3.6	2.2	
	Live	12.9	12.2	10.8	12.2	
	Total	14.4	14.4	14.4	14.4	29 d.f.
6	Dead	1.4	0.0	2.9	1.4	
	Live	12.9	14.4	11.5	12.9	
	Total	14.4	14.4	14.4	14.4	29 d.f.
7	Dead	1.91	0.67	0.90	2.13	
	Live	0.34	1.57	1.34	0.11	
	Total	2.24	2.24	2.24	2.24	3 d.f.

would also apply to comparisons based on other mortality responses. If the extent of heterogeneity differs for the various mortality responses, then separate adjustment factors need be used for each mortality response studied. To get some indication of which situation is the case, we calculate heterogeneity adjustment factors for the responses 21 day mortality conditional on 14 day survival in LeBlanc's Tests A and B. The procedure used for calculating the adjustment factors is very similar to that discussed and illustrated in the previous subsection.

LeBlanc Test A - 21 day mortality conditional on 14 day survival

Group 1: $\hat{p}_{11} = 0.00$, $\hat{p}_{12} = 0.00$, $\hat{p}_{13} = 0.056$, $\hat{p}_{14} = 0.00$, $\hat{\bar{p}}_4 = 0.014$,

$N_{11} = N_{12} = N_{14} = 17$, $N_{13} = 18$, $\bar{N}_1 = 17.25$,

$\hat{Var}(\hat{p}_{ij}) = 0.000784$, $\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1 = 0.00080$

$$\hat{K}_1 = \frac{\hat{Var}(\hat{p}_{1j})}{[\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1]} = \frac{0.000784}{0.000800} = 0.98$$

Group 2: $\hat{p}_{21} = 0.10$, $\hat{p}_{22} = 0.00$, $\hat{p}_{23} = 0.00$, $\hat{p}_{24} = 0.00$, $\hat{\bar{p}}_2 = 0.025$,

$N_{21} = N_{22} = N_{23} = 20$, $N_{24} = 19$, $\bar{N}_2 = 19.75$,

$\hat{Var}(\hat{p}_{2j}) = 0.0025$, $\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2 = 0.00123$

$$\hat{K}_2 = \frac{\hat{Var}(\hat{p}_{2j})}{[\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2]} = \frac{0.0025}{0.00123} = 2.03$$

Group 3: $\hat{p}_{31} = \hat{p}_{32} = \hat{p}_{33} = \hat{p}_{34} = \hat{\bar{p}}_3 = 0$, $N_{31} = 20$, $N_{32} = 19$, $N_{33} = 15$,

$N_{34} = 16$, $\bar{N}_3 = 17.5$, $\hat{Var}(\hat{p}_{3j}) = 0.00$, $\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3 = 0.00$

$$\hat{K}_3 = \frac{\hat{Var}(\hat{p}_{3j})}{[\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3]} = \frac{0.00}{0.00} = \text{indeterminate. Define } \hat{K}_3 \text{ to be 1.00.}$$

Group 4: $\hat{p}_{41} = \hat{p}_{42} = \hat{p}_{44} = 0.00$, $\hat{p}_{43} = 0.111$, $\hat{\bar{p}}_4 = 0.028$, $N_{41} = 17$,

$N_{42} = 20$, $N_{43} = N_{44} = 19$, $\bar{N}_4 = 18.75$, $\hat{Var}(\hat{p}_{4j}) = 0.0031$,

$\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4 = 0.00145$

$$\hat{K}_4 = \frac{\hat{Var}(\hat{p}_{4j})}{[\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4]} = 2.14$$

Group 5: $\hat{p}_{51} = 0.10$, $\hat{p}_{52} = 0.00$, $\hat{p}_{53} = 0.111$, $\hat{p}_{54} = 0.050$, $\hat{\bar{p}}_5 = 0.0065$,

$N_{51} = N_{52} = N_{54} = 20$, $N_{53} = 18$, $\bar{N}_5 = 19.5$,

$\hat{Var}(\hat{p}_{3j}) = 0.0026$, $\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5 = 0.00312$

$$\hat{K}_5 = \frac{\hat{Var}(\hat{p}_{5j})}{[\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5]} = \frac{0.0026}{0.00312} = 0.833$$

Group 6: Delete beaker D from the calculation since it has just one live daphnid on day 14. Thus p_{64} cannot be estimated very precisely.

$\hat{p}_{61} = 0.053$, $\hat{p}_{62} = 0.00$, $\hat{p}_{63} = 0.167$, $\hat{\bar{p}}_6 = 0.073$, $N_{61} = 19$,

$N_{62} = 16$, $N_{63} = 6$, $\bar{N}_6 = 13.67$, $\hat{Var}(\hat{p}_{6j}) = 0.0073$,

$\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6 = 0.0050$

$$\hat{K}_6 = \frac{\hat{Var}(\hat{p}_{6j})}{[\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6]} = \frac{0.0073}{0.0050} = 1.46$$

Group 7: We omit this group from the calculations since the 14 day sample sizes are very small in all 4 beakers (0,0,1,1) and so the p_{7j}'s cannot be estimated very precisely.

An overall inflation factor based on the results from groups 1-6 is

$$\hat{\bar{K}} = \frac{\Sigma_{i=1}^{7} \hat{K}_i}{6} = \frac{0.98 + 2.03 + 1.00 + 2.14 + 0.83 + 1.46}{6} = 1.41$$

Note that $\hat{K}_6$ does not appear to be far removed from the other $\hat{K}_i$'s as it was for the unconditional 21 day mortality. We present a chi square probability plot to determine whether the extent of observed beaker to beaker extrabinomial variation is the same for all groups or is greater for some groups than for others. See the discussion in the previous subsection for a more detailed description of the procedure.

The results for LeBlanc's Test A are summarized below in Table V.5 and are plotted in Figure V.3. The majority of the $\hat{K}_i$'s lie above the cumulative distribution function of a $1/3\ \chi^2_3$ distributed random variable. With the exception of the very large $\hat{K}_i$ for group 6 in Figure V.1, the plots for the unconditional and the conditional mortalities (i.e. Figures V.1 and V.3) look rather similar. In fact, the suggested adjustment factors are $\hat{\bar{K}} = 1.35$ for the unconditional responses (excepting group 6) and $\hat{\bar{K}} = 1.41$ for the conditional responses. In brief, there again appears to be a small degree of extrabinomial variation among beakers within groups but the extent is not too great.

TABLE V.5. LE BLANC TEST A--21 DAY MORTALITY CONDITIONAL ON 14 DAY SURVIVAL--VARIANCE INFLATION FACTOR BY GROUP

Group	Factor (K_i)	Rank	Plotting Position
1	0.98	2	25.00
2	2.03	5	75.00
3	1.00	3	41.67
4	2.14	6	91.67
5	0.83	1	8.33
6	1.46	4	58.33

LeBlanc Test B - 21 Day Mortality Conditional on 14 Day Survival

Group 1: $\hat{p}_{11} = 0.05$, $\hat{p}_{12} = 0.053$, $\hat{p}_{13} = \hat{p}_{14} = 0$, $\hat{\bar{p}}_1 = 0.026$, $N_{11} = 20$,

$N_{12} = 19$, $N_{13} = 20$, $N_{14} = 15$, $\bar{N}_1 = 18.5$,

$\hat{Var}(\hat{p}_{1j}) = 0.00089$, $\hat{\bar{p}}_1(1-\hat{\bar{p}}_1)/\bar{N}_1 = 0.00137$

$$\hat{K}_1 = \frac{0.00089}{0.00137} = 0.65$$

Group 2: $\hat{p}_{21} = 0.056$, $\hat{p}_{22} = \hat{p}_{23} = 0.111$, $\hat{p}_{24} = 0.053$, $\hat{\bar{p}}_2 = 0.083$,

$N_{21} = N_{22} = N_{23} = 18$, $N_{24} = 19$, $\bar{N}_2 = 18.25$,

$\hat{Var}(\hat{p}_{2j}) = 0.00107$, $\hat{\bar{p}}_2(1-\hat{\bar{p}}_2)/\bar{N}_2 = 0.00417$

$$\hat{K}_2 = \frac{0.00107}{0.00417} = 0.26$$

Group 3: $\hat{p}_{31} = \hat{p}_{33} = \hat{p}_{34} = 0.00$, $\hat{p}_{32} = 0.053$, $\hat{\bar{p}}_3 = 0.013$, $N_{31} = 20$, $N_{32} = N_{33} = 19$, $N_{34} = 18$, $\bar{N}_3 = 19$, $\hat{Var}(\hat{p}_{3j}) = 0.00070$, $\hat{\bar{p}}_3(1-\hat{\bar{p}}_3)/\bar{N}_3 = 0.000675$

$$\hat{K}_3 = \frac{0.00070}{0.000675} = 1.04$$

Group 4: $\hat{p}_{41} = 0.150$, $\hat{p}_{42} = \hat{p}_{43} = 0.00$, $\hat{p}_{44} = 0.053$, $\hat{\bar{p}}_4 = 0.051$, $N_{41} = N_{42} = N_{43} = 20$, $N_{44} = 19$, $\bar{N}_4 = 19.75$, $\hat{Var}(\hat{p}_{4j}) = 0.0050$, $\hat{\bar{p}}_4(1-\hat{\bar{p}}_4)/\bar{N}_4 = 0.00245$

$$\hat{K}_4 = \frac{0.0050}{0.00245} = 2.04$$

Group 5: $\hat{p}_{51} = 0.053$, $\hat{p}_{52} = 0.105$, $\hat{p}_{53} = 0.211$, $\hat{p}_{54} = 0.150$, $\hat{\bar{p}}_5 = 0.130$, $N_{51} = N_{52} = N_{53} = 19$, $N_{54} = 20$, $\bar{N}_5 = 19.25$, $\hat{Var}(\hat{p}_{5j}) = 0.00450$, $\hat{\bar{p}}_5(1-\hat{\bar{p}}_5)/\bar{N}_5 = 0.00588$

$$\hat{K}_5 = \frac{0.00450}{0.00588} = 0.77$$

Group 6: $\hat{p}_{61} = \hat{p}_{64} = 0.10$, $\hat{p}_{62} = 0.00$, $\hat{p}_{63} = 0.20$, $\hat{\bar{p}}_6 = 0.10$, $N_{61} = N_{62} = N_{63} = N_{64} = \bar{N}_6 = 20$, $\hat{Var}(\hat{p}_{6j}) = 0.0067$, $\hat{\bar{p}}_6(1-\hat{\bar{p}}_6)/\bar{N}_6 = 0.0045$

$$\hat{K}_6 = \frac{0.0067}{0.0045} = 1.49$$

Group 7: Delete beaker D from the calculation since it has just one live daphnid on day 14. Thus

p_{74} cannot be estimated very precisely.

$\hat{p}_{71} = 0.40$, $\hat{p}_{72} = \hat{p}_{73} = 0.00$, $\hat{\bar{p}}_7 = 0.133$, $N_{71} = 5$, $N_{72} = 14$, $N_{73} = 12$, $\bar{N}_7 = 10.33$, $\hat{Var}(\hat{p}_{7j}) = 0.0533$, $\hat{\bar{p}}_7(1-\hat{\bar{p}}_7)/\bar{N}_7 = 0.0112$

$$\hat{K}_7 = \frac{0.0533}{0.0112} = 4.76$$

The inflation factor from group 7 is somewhat larger than the others. We will check graphically below whether or not it appears to be in line with the others. We present a chi square probability plot to determine whether the extent of extrabinomial variation (if any) appears to be the same for all groups.

The results for LeBlanc's Test B are summarized below in Table V.6 and are plotted in Figure V.4. The $\hat{K}_i$'s lie above the cumulative distribution of a 1/3 χ^2_3 distributed random variable. The factor for group 7 appears to be substantially out of line with those of the other groups. A straight line fitted to the $\hat{K}_i$'s from groups 1 to 6 has a slope 1.36 times that which would be associated with a 1/3 χ^2_3 distributed random variable. Note that the appearance of Figure V.4, based on mortality conditional on 14 day survival, is very similar to that of Figure V.2, based on overall mortality. This suggests that the heterogeneity in group 7 is not just an early life stage phenomenon but persists to later stages of the test.

The appearance of Figure V.4 suggests that we calculate a common inflation factor for groups 1-6 and a separate factor for group 7. The common factor for groups 1-6 can be based on the average of the $\hat{K}_i$'s. Namely

$$\hat{K} = \frac{1}{6} \Sigma_{i=1}^{7} \hat{K}_i = 1.04$$

The factor for group 7 would be $\hat{K}_7 = 4.76$. Note that $\hat{K}$ is slightly smaller than the slope estimated from Figure V.4 (i.e. 1.04 versus 1.36). $\hat{K}$ may be biased downward a bit by the deletion of the largest value from the average. The question of which estimate is better (i.e. average value or slope from probability plot) is a matter for further detailed research and is not pursued further here. For the purpose of definiteness we suggest using $\hat{K}$. Thus we recommend no adjustments in effective sample sizes in groups 1-6 and adjustment by a factor of 4.76 in group 7.

In conclusion, the results in this subsection suggest that different degrees of beaker to beaker heterogeneity hold for different measures of mortality. Therefore separate adjustment factors should be calculated and applied for each measure of mortality studied.

TABLE V.6. LE BLANC TEST B--21 DAY MORTALITY CONDITIONAL ON 14 DAY SURVIVAL--VARIANCE INFLATION FACTOR BY GROUP

Group	Factor ($\hat{K}_i$)	Rank	Plotting Position
1	0.65	2	21.4
2	0.26	1	7.1
3	1.04	4	50.0
4	2.04	6	78.6
5	0.77	3	35.7
6	1.49	5	64.3
7	4.76	7	92.9

D. ADJUSTMENTS TO ACCOUNT FOR BEAKER TO BEAKER HETEROGENEITY OF LENGTH RESPONSES

In Section IV we tested for beaker to beaker heterogeneity within groups for length responses in the data from LeBlanc's Tests A and B. In both tests, statistically significant heterogeneity was found. The tests were carried out by means of a two way nested analysis of variance, the components of which are indicated below. It is assumed in the ANOVA table that there is a balanced situation with I treatment and control groups, J beakers per group, and N daphnids per beaker. This assumption of balance is usually pretty nearly satisfied in aquatic toxicity test length data, except perhaps in those groups which experience high mortality. In particular in LeBlanc's data, all but the highest treatment groups analyzed (i.e. group 6 in Test A and group 7 in Test B) have nearly the same sample sizes.

COMPONENTS OF ANOVA TABLE TO TEST FOR BEAKER TO BEAKER HETEROGENEITY IN LENGTHS

Source	d.f.	Expected Mean Square
Groups	I-1	$\sigma_c^2 + N\,\sigma_b^2 + \frac{JN}{I-1}\,\Sigma_i\,\alpha_i^2$
Beakers Within Groups	I(J-1)	$\sigma_c^2 + N\,\sigma_b^2$
Daphnids Within Groups	IJ(N-1)	σ_c^2

In agreement with the notation used in Section IV, α_i represents the fixed group effect, σ_b^2 represents the variance of the random beaker effect within groups, and σ_c^2 represents the variance of the random daphnid effect within beakers.

It is clear from the ANOVA table that the mean square for beakers within groups provides a correct error yardstick for inferences about group effects, whether or not beaker to beaker heterogeneity exists. In fact basing inferences on this error term is equivalent to analyzing the length data on a per beaker basis (i.e. analyzing only average lengths within beakers). This is currently the most commonly used approach for the analysis of such data. This is a conservative approach.

However, although use of the beakers within groups mean square is correct, it is based on relatively few degrees of freedom (I(J-1)) and thus can lead to reduced sensitivity of inferences about lengths if I(J-1) is small. This would be the case particularly if J was 2 or 3. It would be beneficial for the sensitivity of the analyses if the number of degrees of freedom available for the error yardstick could be increased, perhaps by somehow combining information from that in another mean square. A scheme for doing this is discussed below.

The usual approach to combining information from several mean squares in analysis of variance is based on a preliminary test. Namely first test the hypothesis that $\sigma_b^2 = 0$ by comparing the mean square for beakers within groups to the mean square for daphnids within beakers. If the hypothesis is rejected, use the beakers within groups mean square with I(J-1) degrees of freedom as an error yardstick. If the hypothesis is not rejected, then combine the two sums of squares for σ_c^2 and use a pooled error estimate based on I(J-1) + IJ(N-1) = I(JN-1) degrees of freedom. This procedure corresponds to either carrying out comparison of lengths across groups on a per beaker basis or else pooling data across beakers within groups and carrying out comparisons on a per daphnid basis, ignoring the existence of replicate beakers. This approach is somewhat dichotomous. One of two somewhat different procedures is carried out, depending on whether the preliminary test rejects or accepts. It would be desirable to use a procedure which provides a continuum of options between the above two extremes and that does not rely on the outcome of a preliminary test. We now describe such a procedure.

Each individual length has variance $\sigma_c^2 + \sigma_b^2$. Since the lengths from the same beaker are correlated due to the beaker effects, the average over all NJ lengths within a group has variance $(\sigma_c^2 + N\,\sigma_b^2)/NJ \equiv \sigma_c^2/NJ + \sigma_b^2/J$. Suppose we wish to account for the within beaker correlation by reducing the effective sample size within beakers from N to x and then treating the "adjusted samples" as if they were independent, with variance $\sigma_c^2 + \sigma_b^2$. We would thus disregard beakers and carry out analyses on a per daphnid basis. To determine the effective sample size, x, per beaker we equate the variances of sample averages under the true and hypothecated situations. Namely

$$\frac{\sigma_c^2 + \sigma_b^2}{Jx} = \frac{\sigma_c^2 + N\,\sigma_b^2}{JN}$$

Thus

$$x = \frac{N(\sigma_c^2 + \sigma_b^2)}{\sigma_c^2 + N\,\sigma_b^2} \quad \text{and so } 1 \leq x \leq N$$

We treat the data as if there were Jx independent observations per group. The degrees of freedom to estimate error would then be I(Jx-1). Now

$$I(Jx-1) = I\left(JN \frac{\sigma_c^2 + \sigma_b^2}{\sigma_c^2 + N\,\sigma_b^2} - 1\right) \quad .$$

It is easy to see that

$$I(J-1) \leq I(Jx-1) \leq I(JN-1)$$

the extremes occurring as $\sigma_b^2 \to \infty$ or $\sigma_b^2 \to 0$. Thus using I(Jx-1) degrees of freedom is a compromise between using I(J-1) d.f. and I(JN-1) d.f.

In order to calculate x we need to know σ_b^2 and σ_c^2. In general these quantities are unknown, however they can be estimated from the data. Let

MSI = mean square for beakers within groups.
MSE = mean square for daphnids within beakers.

Then

$$\hat{\sigma}_c^2 = \text{MSE}, \quad \widehat{\sigma_c^2 + N\,\sigma_b^2} = \text{MSI}, \quad \widehat{\sigma_c^2 + \sigma_b^2} = \frac{1}{N}\,\text{MSI} + \left(1 - \frac{1}{N}\right)\text{MSE} \quad .$$

Thus

$$\hat{x} = \frac{\text{MSI} + (N-1)\,\text{MSE}}{\text{MSI}} = 1 + (N-1)\,\text{MSE/MSI} \quad .$$

We constrain $\hat{x}$ to be bounded by N. Namely we take

$$\hat{x} = \min\left[1 + (N-1)\,\text{MSE/MSI},\ N\right] \quad .$$

Since there is uncertainty in $\hat{x}$, a very conservative assumption would be to take x to be its minimum value of 1. This would correspond to using MSI as an error yardstick with I(J-1) d.f. Thus there is no pooling of information from MSE. A less conservative assumption would be to use $I(J\hat{x}-1)$ d.f. We use max(MSI,MSE) as the error yardstick, but utilize the information from MSE to increase the degrees of freedom assumed from I(J-1) to $I(J\hat{x}-1)$. A compromise between these two values of x would be to calculate a lower confidence bound for x and use this value in the expression for degrees of freedom. Let $\underset{\sim}{x}$ denote a 95 percent lower confidence bound on x. We use $I(J\underset{\sim}{x}-1)$ degrees of freedom in conjunction with max(MSI,MSE).

A lower confidence bound on x corresponds to an upper confidence bound on σ_b^2/σ_c^2 since

$$x = N(1 + \sigma_b^2/\sigma_c^2)/(1 + N\,\sigma_b^2/\sigma_c^2) = 1 + (N-1)/(1 + N\,\sigma_b^2/\sigma_c^2) \quad .$$

Now

$$\text{MSI/MSE} \sim (1 + N\,\sigma_b^2/\sigma_c^2)\,F_{I(J-1),\ IJ(N-1)} \quad .$$

Thus

$$\frac{MSI/MSE}{F^{(0.05)}_{I(J-1),\ IJ(N-1)}} = (MSI/MSE)\ F^{(0.95)}_{IJ(N-1),\ I(J-1)}$$

is an upper 95 percent confidence bound on $(1 + N\sigma_b^2/\sigma_c^2)$. Therefore

$$\underset{\sim}{x} = \min\left\{1 + (N-1)/[MSI/MSE]\ F^{(0.95)}_{IJ(N-1)\ \ I(J-1)},\ N\right\}$$

The suggested compromise procedure is to use max[MSI,MSE] as the error yardstick but to increase the effective degrees of freedom to $I(J\underset{\sim}{x}-1)$. In subsequent comparisons among treatment means we use $(\max[MSI,MSE]/JN)^{1/2}$ as the estimated standard error of treatment group averages and we associate $I(J\underset{\sim}{x}-1)$ d.f. with it.

This procedure has intuitive appeal as means of increasing the sensitivity of inferences about lengths when I(J-1) is small. Its precise theoretical properties need to be investigated in greater detail, perhaps by a Monte Carlo study.

We now apply this procedure to the length data from LeBlanc's Tests A and B. Although N is not entirely constant across beakers, especially in the highest treatment groups, we use the above expressions with an average value of N.

LeBlanc Test A

From Section IV, MSI = 0.2722, MSE = 0.118, I=6, J=4, $N \equiv \bar{N} = 401/24 = 16.71$. Thus IJ(N-1) = 377, I(J-1) = 18, $F_{377,\ 18}(0.95) = 1.93$, MSI/MSE = 2.435. Therefore

$$\underset{\sim}{x} = 1 + (15.71)/[2.345)(1.93)] = 4.34$$

$$\hat{x} = 1 + (15.75)/2.435 = 7.47$$

We thus see that the point estimate of the effective number of daphnids per beaker is 7.5 and a 95 percent lower confidence bound on this is 4.3. We associate $I(J\underset{\sim}{x}-1) = 6((4)(4.34)-1) = 98.2 \simeq 98$ d.f. with the mean square for beakers within groups. This compares with 6(4-1) = 18 d.f. associated with this mean square if we pool no information from the within beakers mean square.

LeBlanc Test B

From Section IV, MSI = 0.2707, MSE = 0.1387, I=7, J=4, $N \equiv \bar{N} = 459/28 = 16.39$. Thus IJ(N-1) = 431, I(J-1) = 21, $F_{431,\ 21}(0.95) = 1.84$, MSI/MSE = 1.952. Therefore

$$\underset{\sim}{x} = 1 + (15.39/[(1.95)(1.84)] = 5.28$$

$$\hat{x} = 1 + (15.39)/1.952 = 8.88$$

The point estimate of the effective number of daphnids per beaker is 8.9 and a 95 percent lower confidence bound on this is 5.3. We associate $I(Jx-1) = 7((4)(5.28)-1) = 140.84 \simeq 141$ d.f. with the mean square for beakers within groups. This compares with $7(4-1) = 21$ d.f. associated with this mean square if we pool no information from the within beakers mean square.

The model upon which the adjustments in degrees of freedom in LeBlanc's Tests A and B was based (i.e. upon which the within and between mean squares were effectively pooled) assumes that the components of variation are the same across groups. Namely it is assumed that there is random beaker to beaker variation with variance σ_b^2 and there is random within beaker variation with variance σ_c^2. Before applying adjustments based on this model one should determine whether all the groups appear to have the same variability or whether there are one or two groups with variability substantially higher than the remainder, that inflate the overall estimates of variability for the other groups. If that were the case, separate variability estimates would need to be calculated for the group or groups with large variability and subsequent analyses would need to take this into account, perhaps by using weighted least squares.

We discuss below a graphical procedure for determining whether the observed beaker to beaker variation is due to all groups or whether it is due just to one or two and we apply this procedure to the length responses from LeBlanc's Tests A and B. The procedure consists of calculating the mean square for beakers within groups separately within each group and normalizing these values by the pooled mean square for daphnids within beakers. Under the assumption of common variance structure across groups, these normalized ratios estimate $1 + N\sigma_b^2/\sigma_c^2$ and are distributed approximately as $(1 + N\sigma_b^2/\sigma_c^2)\ \chi^2_{J-1}/(J-1)$. We should thus see a straight line or at least a smooth curve when the ordered normalized ratios are plotted on chi square probability paper. If one or two points are far removed from the others, this suggests differing variability structures across groups and subsequent analyses would need to reflect this.

The results for LeBlanc's Test A are summarized below. In this test J=4 and I=6. The rank is just the order of the ratios, from smallest to largest. The plotting position for the ratio with rank i is $100\ (i-0.5)/6$.

TABLE V.7. LE BLANC TEST A--NORMALIZED MEAN SQUARE RATIOS BY GROUP

Group	Normalized Ratio	Rank	Plotting Position
1	3.678	5	75.00
2	0.579	2	25.00
3	5.757	6	91.67
4	1.680	3	41.67
5	2.405	4	58.33
6	0.509	1	8.33

The normalized ratios are plotted versus their plotting positions on chi square probability paper with 3 d.f. The results are shown in Figure V.5. The reference line in that plot corresponds to the theoretical c.d.f. of a random variable with distribution $1/3\ \chi^2_3$. The points are seen to fall on a straight line with slope substantially in excess of that which would be expected if there were no beaker to beaker variation. We conclude that this plot suggests the presence of beaker to beaker variation in lengths and that the variation structure is constant across groups.

The results for LeBlanc's Test B are summarized below. In this test J=4 and I=7. The terms rank and plotting position have the same meaning as for Test A.

TABLE V.8. LE BLANC TEST B--NORMALIZED MEAN SQUARE RATIOS BY GROUP

Group	Normalized Ratio	Rank	Plotting Position
1	0.392	1	7.14
2	3.946	7	92.86
3	3.876	6	78.57
4	0.935	4	50.00
5	0.451	2	21.43
6	3.421	5	64.29
7	0.639	3	35.71

The normalized ratios are plotted versus their plotting positions on chi square probability paper with 3 d.f. The results are shown in Figure V.6. The reference line in that plot corresponds to the theoretical c.d.f. of a random variable with distribution $1/3\ \chi^2_3$. The ratios seem to fall into two distinct subsets. The lower four values appear to be compatible with the absence of beaker to beaker variation in their groups. The upper three values are separated from the others and indicate the presence of beaker to beaker variation. This graph should be discussed with the investigator to determine if any identifiable experimental factors could explain the apparent dichotomy between the variability in groups 1, 4, 5 and 7 on the one hand and that in groups 2, 3 and 6 on the other. If so, separate variability estimates might be used for these two subsets of groups. In the absence of such information we will use the overall adjustment results derived earlier in this subsection.

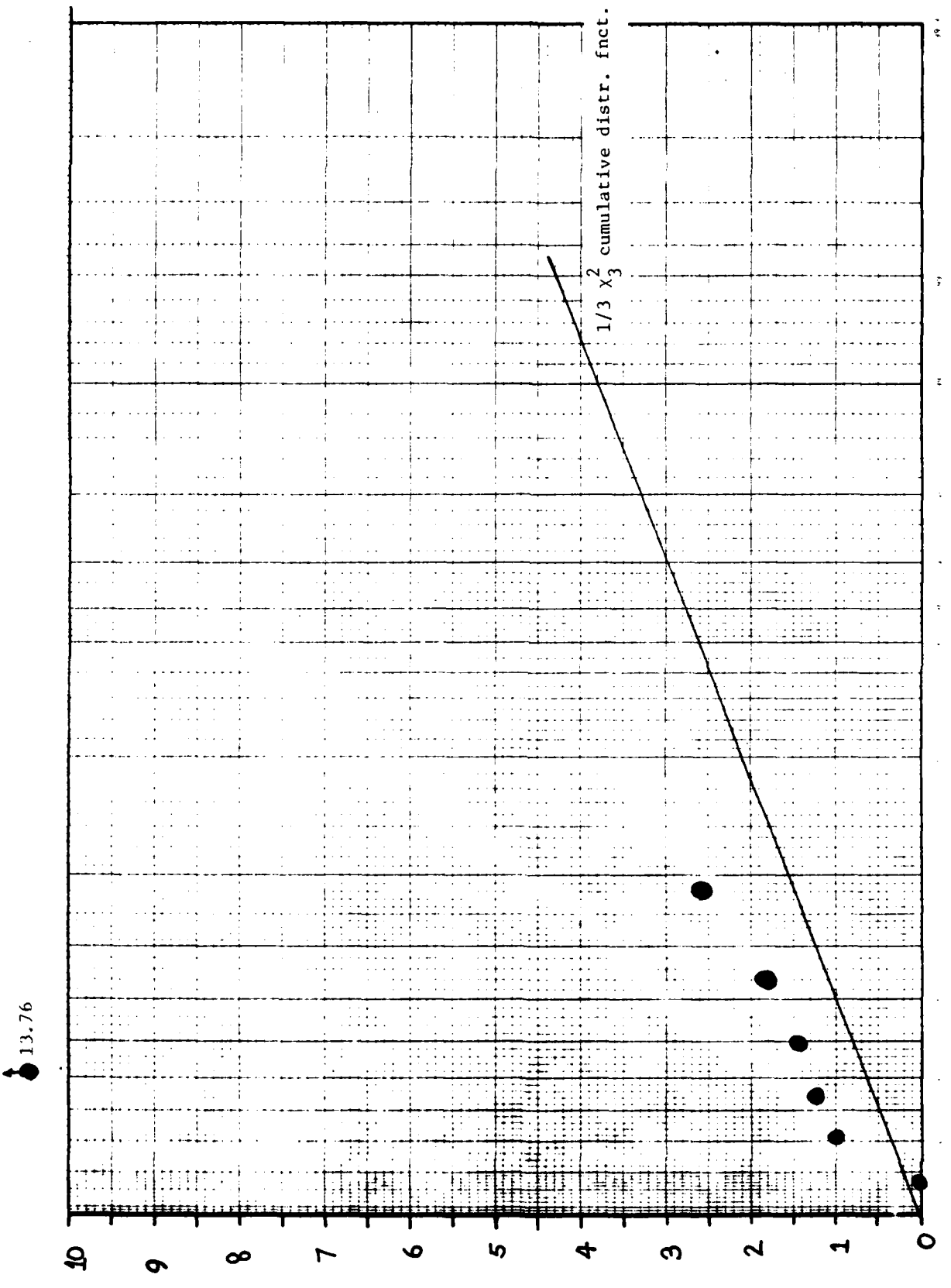


Figure V.1. LeBlanc Test A. Chi square probability plot of mortality data variance inflation factors.

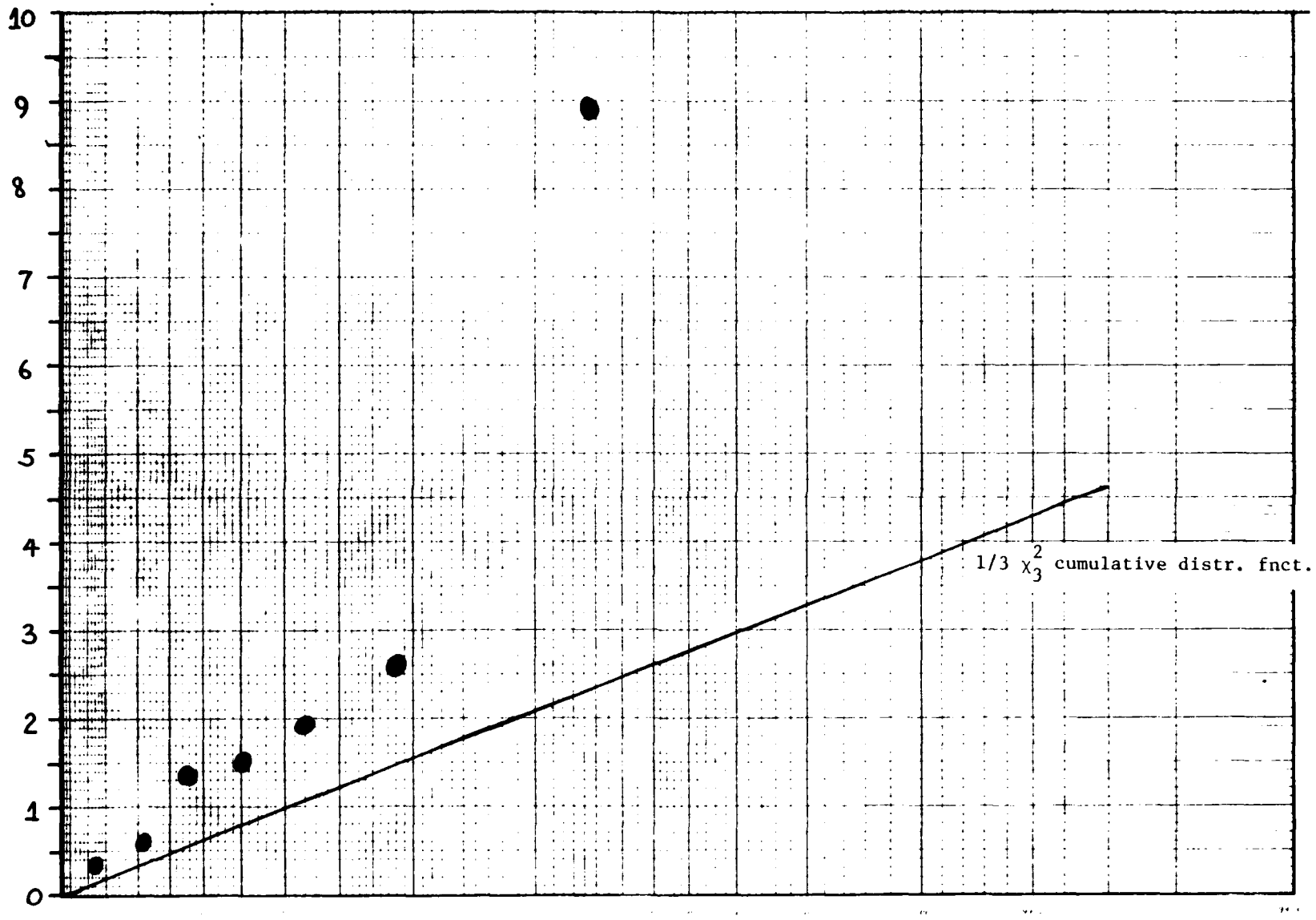


Figure V.2. LeBlanc Test B. Chi square probability plot of mortality data variance inflation factors.

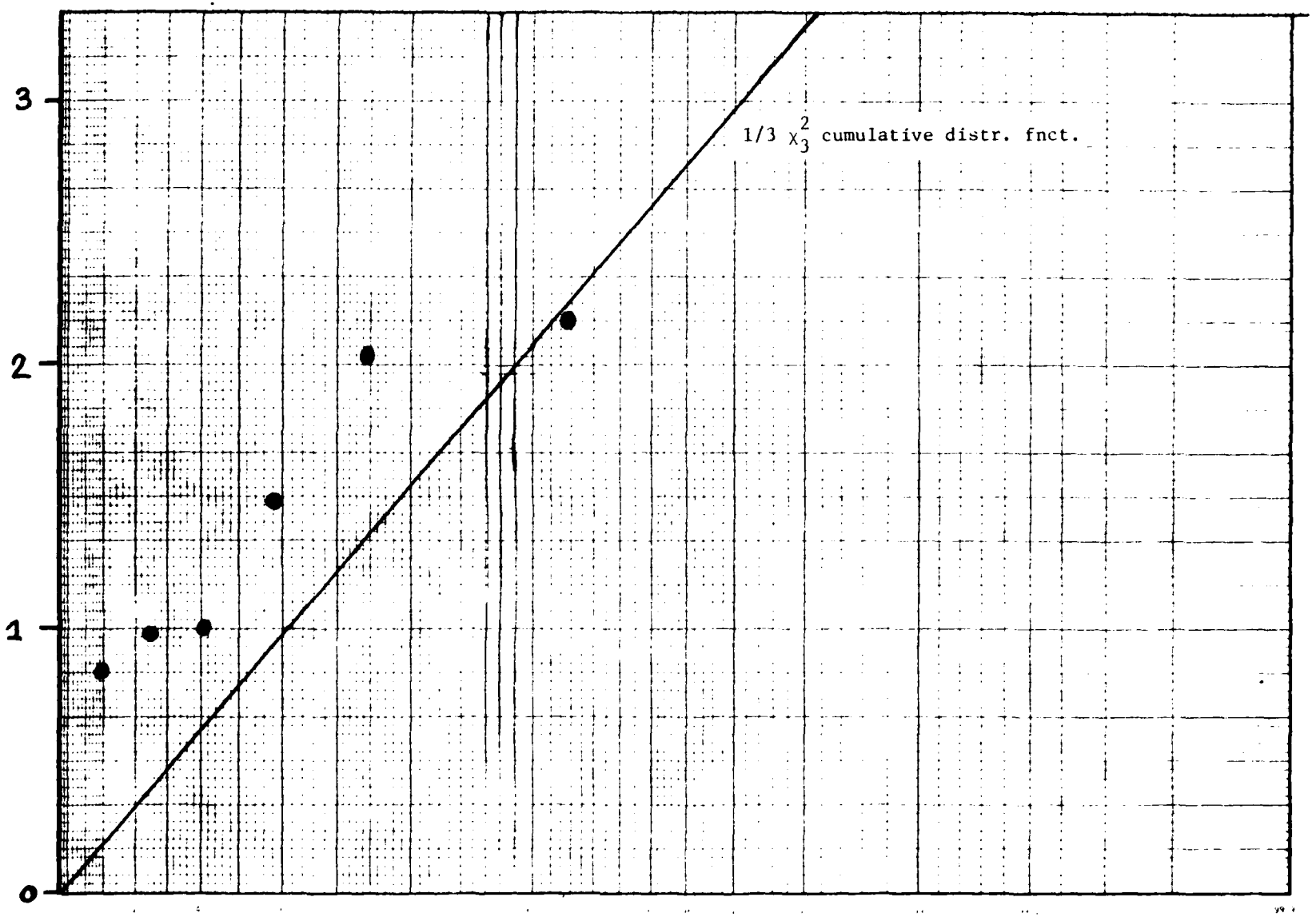


Figure V.3. LeBlanc Test A. 21 day mortality conditional on 14 day survival. Chi square probability plot of variance inflation factors.

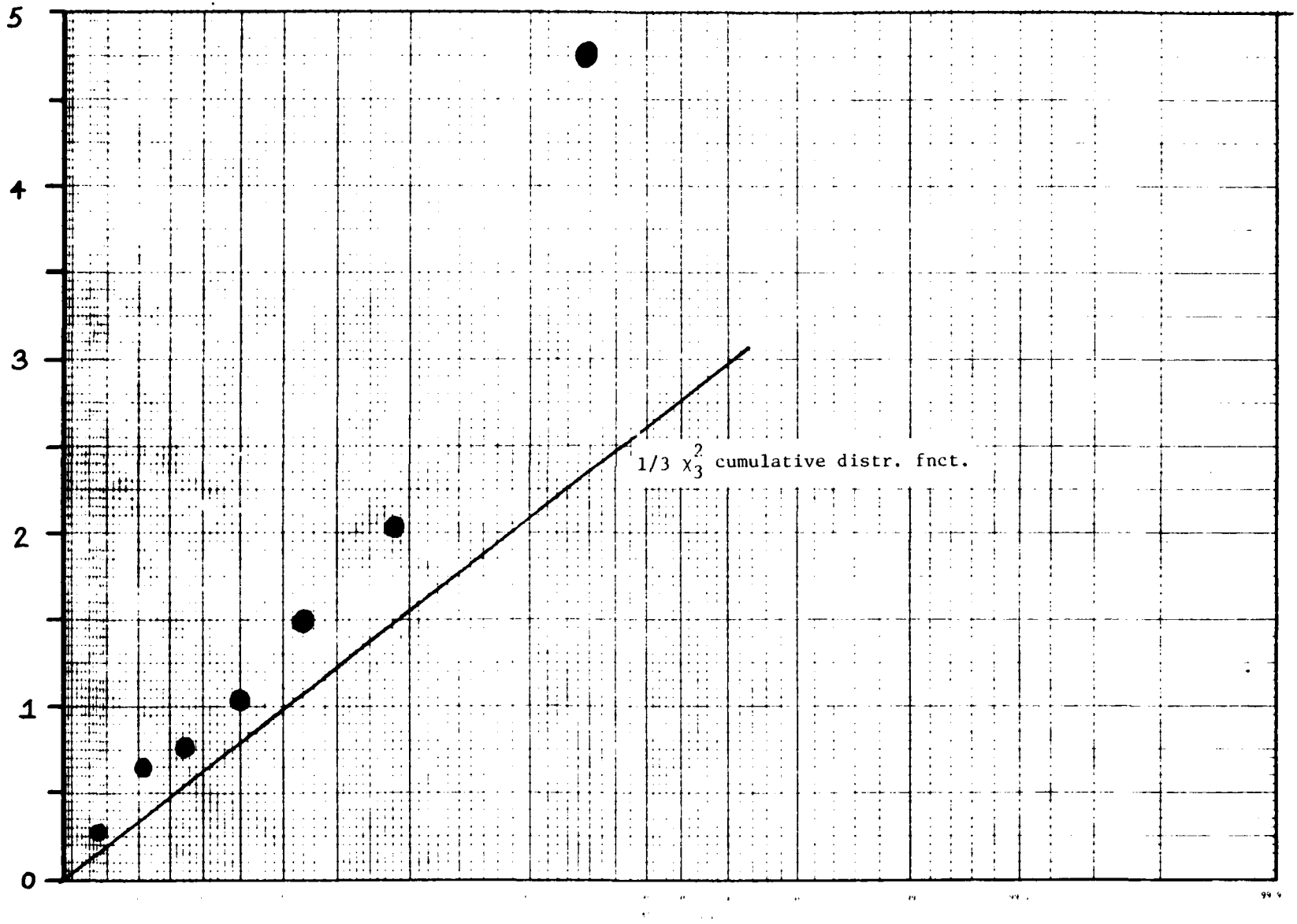


Figure V.4. LeBlanc Test B. 21 day mortality conditional on 14 day survival. Chi square probability plot of variance inflation factors.

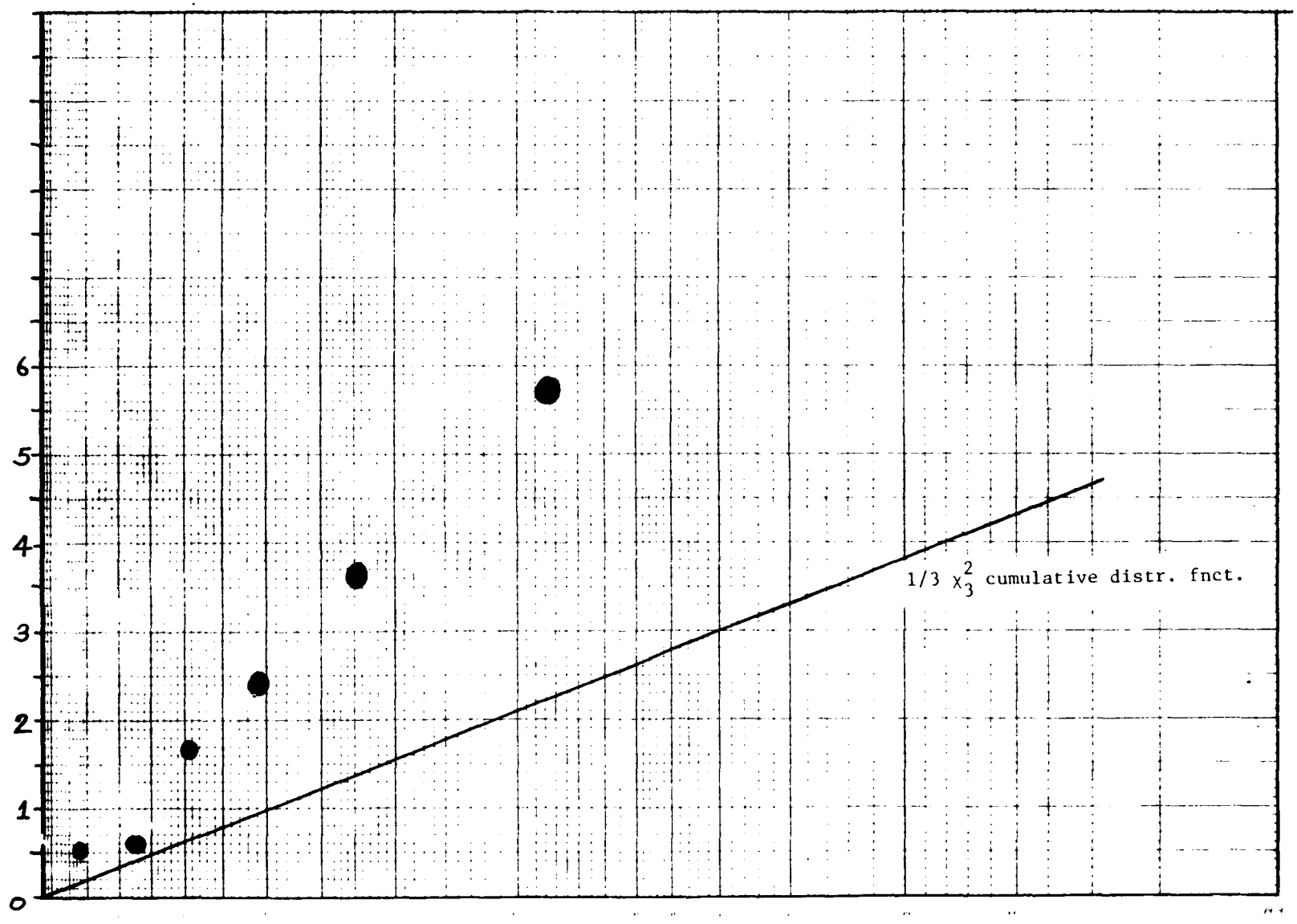


Figure V.5. LeBlanc Test A. Chi square probability plot of mean squares for beakers within groups normalized by mean square for daphnids within beakers--lengths.

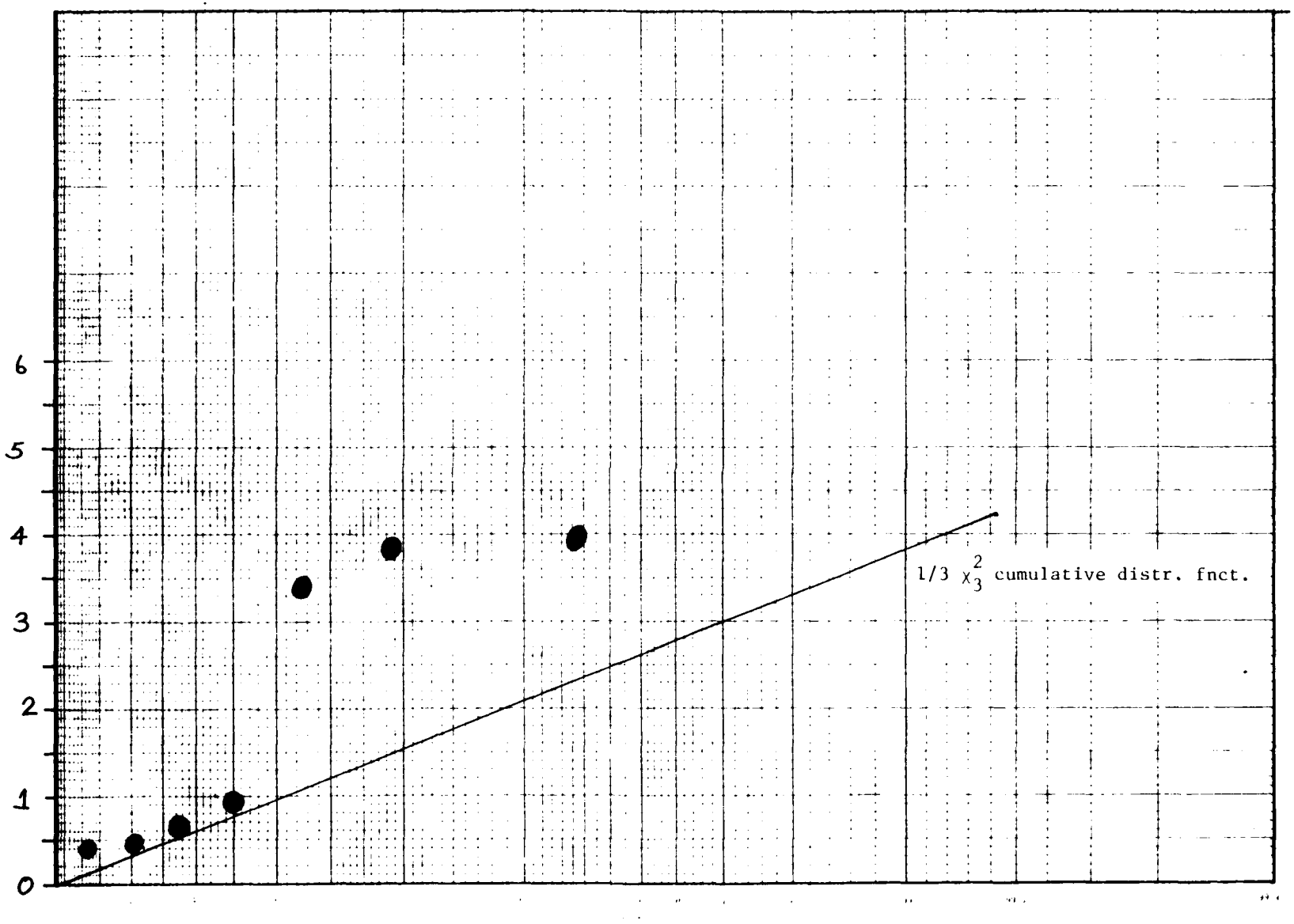


Figure V.6. LeBlanc Test B. Chi square probability plot of mean squares for beakers within groups normalized by mean square for daphnids within beakers--lengths.

VI. OUTLIER DETECTION PROCEDURES

A. BACKGROUND

Another preliminary analysis of importance is the detection of responses which do not appear to be in conformance with the substantial majority of responses. Such exceptional responses are often referred to as "outliers". Outlier detection procedures are used to decide how extreme a response must be in order to rule out the possibility that its value is reasonably likely to be due just to random variation. Feder and Collins [1], Section X and Subsection XVII B, discuss outlier detection procedures for the analysis of mortality and length data from toxicity tests on fathead minnows. The discussion here is patterned after that.

Before discussing the details of outlier detection procedures for the various responses considered here, we need to discuss an important conceptual issue. Outlier detection procedures look for responses which differ from a priori comparable responses more than would be expected based on random variation. In the previous section we estimated the extent of beaker to beaker variation and adjusted for this variation by utilizing the differences in responses observed in a priori comparable beakers (i.e. within the same treatment groups). The presence of outliers will inflate the estimates of beaker to beaker random variation. Conversely, the presence of random beaker to beaker variation might cause extreme but naturally occurring responses to appear as outliers. Thus the notions of outliers and of inflated beaker to beaker random variation are somewhat confounded and obscure one another. If an individual observation or a beaker average looks extreme there is no way to tell, based on the data alone, whether that response represents natural variation or whether it comes from a separate population, due perhaps to some deviation in biological material or experimental technique. This is a matter for judgement on the part of the investigator. Statistical methods can point to the extreme or out of line responses. They cannot determine the reasons for this behavior or whether the observations in question should be retained, deleted, or discounted.

Outlier detection procedures and beaker to beaker heterogeneity adjustment procedures impact on one another. Consider the mortality responses for LeBlanc's Tests A and B plotted in Figures II.1 and II.3 and for Goulden's isophorone test plotted in Figure II.8. The responses in group 6 of LeBlanc's Test A, group 7 of LeBlanc's Test B, and group 5 of Goulden's isophorone test appear to be widely separated from one another. Is the separation due to natural random variation or are one or more beakers out of line from the others? Based on the appearances of the plots, the differences among these beakers in the LeBlanc tests appear to be due to random variation. Two of the four beakers show relatively low mortality while the other two show relatively high mortality. None of the responses are out of line with those from other groups. The situation in group 5 of Goulden's survival data is a bit different. The beaker with 100 percent mortality appears to be out of line with the other two multiple daphnid beakers and with the seven individual daphnid beakers in its group. This

suggests that it may be a possible outlier. However it must be stressed that these interpretations are subjective and may differ among investigators or among data analysts.

Perhaps as a rule of thumb, extreme deviations should be regarded as evidence of random beaker to beaker variation unless there is specific evidence to the contrary. This is a matter for further discussion.

We adopt the following order of analysis. Based on the appearance of the preliminary scatterplots, any responses that are out of line and whose deviations can be traced back to faulty biological material, departures in procedure from test protocol, accidents, etc will be screened out. The remaining responses will be tested for beaker to beaker heterogeneity and appropriate error yardsticks will be calculated or data adjustments made. Outlier detection tests will then be carried out incorporating the previously calculated error yardsticks and data adjustments.

We illustrate the outlier detection procedures below.

B. OUTLIER DETECTION PROCEDURES APPLIED TO MORTALITY DATA

The procedure followed is similar to that discussed in Feder and Collins [1], Section X. Consider the i-th group. Let (X_{i1}, N_{i1}), $X_{i2}, N_{i2}),\ldots,$ (X_{iJi}, N_{iJi}) denote the numbers of responses and the number of daphnids in the various beakers $j=1,\ldots, J_i$. Delete the subscripts i in subsequent discussion, for ease of notation. Note that (X_j, N_j) denote the effective numbers of responses and daphnids respectively, after adjusting for beaker to beaker heterogeneity, rather than the original numbers. The arc sin variance stabilizing transformation is first carried out. In particular let $\hat{p}_j$, $\hat{p}$ denote the response probability estimates for the j-th beaker and for the group respectively. It can be shown that $2N_j^{1/2}(1-N_j/N)^{-1/2}$ [arc sin $(\hat{p}_j^{1/2})$ - arc sin $(\hat{p}^{1/2})$], $j=1,\ldots, J$ have approximate standard normal distributions as the N_j's approach ∞. Graphical and numerical outlier detection procedures are based on these standardized values. For formal inferences we account approximately for the correlations among the standardized values within groups (approximately $-1/(J-1)$ if the N_j's are about equal) by treating the J standardized values within each group as if they were J-1 independent values. This of course has the most effect when J=2.

We now apply these transformations to construct graphical outlier detection displays based on normal probability plotting and associated formal outlier detection tests.

LeBlanc Test A

I=7, J=4 (i.e. 7 groups, 4 beakers per group)

	X_j[1]	N_j[1]	$\hat{p}_j$	$\hat{p}$	N_j/N	$N_j\hat{p}$	Q^2
Group 1							
Beaker A	2.2	14.8	0.149	0.149	0.25	2.205	0.0
Beaker B	2.2	14.8	0.149	0.149	0.25	2.205	0.0
Beaker C	2.2	14.8	0.149	0.149	0.25	2.205	0.0
Beaker D	2.2	14.8	0.149	0.149	0.25	2.205	0.0
Group 2							
Beaker A	1.5	14.8	0.101	0.037	0.25	0.548	1.154
Beaker B	0.0	14.8	0.0	0.037	0.25	0.548	-1.720
Beaker C	0.0	14.8	0.0	0.037	0.25	0.548	-1.720
Beaker D	0.7	14.8	0.047	0.037	0.25	0.548	0.222
Group 3							
Beaker A	0.0	14.8	0.0	0.125	0.25	1.850	-3.211
Beaker B	0.7	14.8	0.047	0.125	0.25	1.850	-1.269
Beaker C	3.7	14.8	0.250	0.125	0.25	1.850	1.441
Beaker D	3.0	14.8	0.203	0.125	0.25	1.850	0.942
Group 4							
Beaker A	2.2	14.8	0.149	0.086	0.25	1.273	0.877
Beaker B	0.0	14.8	0.0	0.086	0.25	1.273	-2.644
Beaker C	2.2	14.8	0.149	0.086	0.25	1.273	0.877
Beaker D	0.7	14.8	0.047	0.086	0.25	1.273	-0.703
Group 5							
Beaker A	1.5	14.8	0.101	0.088	0.25	1.302	0.198
Beaker B	0.0	14.8	0.0	0.088	0.25	1.302	-2.676
Beaker C	3.0	14.8	0.203	0.088	0.25	1.302	1.477
Beaker D	0.7	14.8	0.047	0.088	0.25	1.302	-0.734
Group 6							
Beaker A	0.14	1.45	0.097	0.500	0.25	0.725	-1.303
Beaker B	0.29	1.45	0.200	0.500	0.25	0.725	-0.895
Beaker C	1.09	1.45	0.752	0.500	0.25	0.725	0.734
Beaker D	1.38	1.45	0.952	0.500	0.25	0.725	1.570

	X_j[1]	N_j[1]	$\hat{p}_j$	$\hat{p}$	N_j/N	$N_j\hat{p}$	Q[2]
Group 7							
Beaker A	14.8	14.8	1.000	0.988	0.25	14.622	0.975
Beaker B	14.8	14.8	1.000	0.988	0.25	14.622	0.975
Beaker C	14.1	14.8	0.953	0.988	0.25	14.622	-0.966
Beaker D	14.8	14.8	1.00	0.988	0.25	14.622	0.975

To prepare the normal probability plot we order the standardized values and plot the i-th smallest against the plotting position 100 x (i-0.5)/28 on the probability scale. These values are indicated below.

i	1	2	3	4	5	6	7	8
Ordered Value	-3.211	-2.676	-2.644	-1.720	-1.720	-1.303	-1.269	-0.966
Plotting Position	1.8	5.4	8.9	12.5	16.1	19.6	23.2	26.8

i	9	10	11	12	13	14	15	16
Ordered Value	-0.895	-0.734	-0.703	0.0	0.0	0.0	0.0	0.198
Plotting Position	30.4	33.9	37.5	41.1	44.6	48.2	51.8	55.4

i	17	18	19	20	21	22	23	24
Ordered Value	0.222	0.734	0.877	0.877	0.942	0.975	0.975	0.975
Plotting Position	58.9	62.5	66.1	69.6	73.2	76.8	80.4	85.9

i	25	26	27	28
Ordered Value	1.154	1.441	1.477	1.570
Plotting Position	87.5	91.1	94.6	98.2

[1] X_j, N_j represent effective values, after adjustment for beaker to beaker heterogeneity.

[2] $Q = (1-N_j/N)^{-1/2}\ 2\sqrt{N_j}\ [\arcsin\sqrt{\hat{p}_j} - \arcsin\sqrt{\hat{p}}]$

The normal probability plot of these points is shown in Figure VI.1. A theoretical N(0,1) distribution line is drawn in for reference. It is evident that the standardized values lie on a smooth curve which is definitely not normal. The lack of normality is undoubtedly due to the relatively small (effective) sample sizes per beaker coupled with the relatively extreme average mortality rates within each group (i.e. close to 0 or to 1). Note that the minimum expected frequencies min $(N_j\hat{p}, N_j\hat{q})$ within each group (based on effective sample sizes) are less than 2 in all but the first group where the minimum expected frequency is 2.2. In fact, three of the seven groups have minimum expected frequencies less than 1. The validity of the normal approximation to arc sin $(p^{1/2})$ is questionable for such small expected frequencies. Furthermore the five smallest standardized values correspond to observed frequencies of 0 deaths per beaker, where the normal approximation is least reasonable.

Despite the lack of normality the smooth and continuous nature of the curve in Figure VI.1 suggests no evidence of outlying responses. The extreme discrepancies among mortality rates observed in group 6 have been attributed to random beaker to beaker variation and have been accounted for by adjusting the effective sample sizes downward very extremely. Thus the standardized values from these groups do not appear to be outliers. Without such prior adjustment they undoubtedly would have.

It should be noted that confirmatory tests should be carried out before declaring a beaker response to be an outlier, based just on the appearance of the normal probability plot. This is especially true when the response corresponds to small observed frequencies (especially 0) where the normal approximation is most questionable. Such confirmatory tests for mortality responses are discussed in Feder and Collins [1], Section X. In particular, an exact confirmatory test, based on Poisson theory, is discussed for the case when average group response rates are less than 0.1 or greater than 0.9. This is the case in four of the seven groups in the Test A mortality data and two of the other three groups are close. In other cases, chi square tests of homogeneity can be carried out, with adjustments for small expected frequencies or based on exact small sample theory (e.g. Fisher's exact test).

In brief, we conclude that there is no evidence of outlying mortality responses in this data set.

We now consider some of the other data sets. The pattern of the mortality responses in LeBlanc's Test B is very nearly the same as that in Test A, which we have just discussed in detail. We thus omit discussion of outlier detection in Test B. The results will undoubtedly be very similar to those from Test A--namely no outliers. Since Chapman's test consists of just one daphnid per beaker we cannot compare mortality response rates among beakers within groups in this test. The mortality patterns in Adams' test with selenium are quite consistent among beakers within groups. The first four groups have essentially no mortality in any beaker while the last four groups have 100 percent mortality in each beaker. (See Figure II.5.) There is thus no suggestion of any outlier responses in this test and so we omit discussion of outlier detection procedures.

The situation with Goulden's test with isophorone is a bit different. (See Figure II.8. Note that survival rates are plotted in this figure rather than mortality rates.) The beaker in group 5 with 100 percent mortality appears to be out of line with the others in that group. This extreme response may be due to random variation or to some identifiable cause that would make it an outlier. There is no way to distinguish between these two possibilities solely on the basis of the data. The test records must be thoroughly reviewed and biological judgement must be brought to bear. However we note that there is just one beaker apparently out of line with the others and there is no overall evidence of beaker to beaker heterogeneity within groups among the beakers with multiple daphnids (see Section III). We will thus treat this response as a potential outlier and carry out graphical and analytical outlier detection procedures to confirm or refute this conjecture. We could have alternatively adjusted the effective sample sizes downward in all the beakers in group 5 to reflect random beaker to beaker variation in this group only. The decision to carry out the outlier detection procedure rather than the heterogeneity adjustment procedure is somewhat subjective.

Goulden - Isophorone

Since there is a suggestion of greater mortality among the multiply housed daphnids (see Section III) we compare mortality rates only among the beakers with multiple daphnids.

I=6, J=3 (i.e. 6 groups, 3 beakers per group)

	X_j[1]	N_j[1]	$\hat{p}_j$	$\hat{p}$	N_j/N	$N_j\hat{p}$	Q[2]
Group 1							
Beaker A	1	5	0.20	0.133	0.333	0.667	0.495
Beaker B	1	5	0.20	0.133	0.333	0.667	0.495
Beaker C	0	5	0.00	0.133	0.333	0.667	-2.045
Group 2							
Beaker A	1	5	0.20	0.067	0.333	0.333	1.105
Beaker B	0	5	0.00	0.067	0.333	0.333	-1.434
Beaker C	0	5	0.00	0.067	0.333	0.333	-1.434
Group 3							
Beaker A	1	5	0.20	0.067	0.333	0.333	1.105
Beaker B	0	5	0.00	0.067	0.333	0.333	-1.434
Beaker C	0	5	0.00	0.067	0.333	0.333	-1.434

	X_j[1]	N_j[1]	$\hat{p}_j$	$\hat{p}$	N_j/N	$N_j\hat{p}$	Q[2]
Group 4							
Beaker A	2	5	0.40	0.200	0.333	1.000	1.211
Beaker B	1	5	0.20	0.200	0.333	1.000	0.000
Beaker C	0	5	0.00	0.200	0.333	1.000	-2.540
Group 5							
Beaker A	5	5	1.00	0.533	0.333	2.667	4.121
Beaker B	2	5	0.40	0.533	0.333	2.667	-0.732
Beaker C	1	5	0.20	0.533	0.333	2.667	-1.943
Group 6							
Beaker A	4	5	0.80	0.867	0.333	4.333	-0.495
Beaker B	4	5	0.80	0.867	0.333	4.333	-0.495
Beaker C	5	5	1.00	0.867	0.333	4.333	2.045

To prepare the normal probability plot we order the standardized values and plot the i-th smallest against the plotting position 100 x (i - 0.5)/18 on the probability scale. These values are indicated below.

i	1	2	3	4	5	6
Ordered Value	-2.540	-2.045	-1.943	-1.434	-1.434	-1.434
Plotting Position	2.8	8.3	13.9	19.4	25.0	30.6

i	7	8	9	10	11	12
Ordered Value	-1.434	-0.732	-0.495	-0.495	0.00	0.495
Plotting Position	36.1	41.7	47.2	52.8	58.3	63.9

i	13	14	15	16	17	18
Ordered Value	0.495	1.105	1.105	1.211	2.045	4.121
Plotting Position	69.4	75.0	80.6	86.1	91.7	97.2

[1] X_j, N_j represent actually observed values, with no adjustments for beaker to beaker heterogeneity.

[2] $Q = (1-N_j/N)^{-1/2}\ 2\ \sqrt{N_j}\ [\arcsin \sqrt{\hat{p}_j} - \arcsin \sqrt{\hat{p}}]$

The normal probability plot of these points is shown in Figure VI.2. A theoretical N(0,1) distribution line is drawn in for reference. All but the largest standardized values lie approximately on a straight line with slope slightly greater than that based on the assumptions of no beaker to beaker heterogeneity and no outliers (slope is about 1.5 as compared with 1.0 for the theoretical N(0,1) line). The departure from the standard normal distribution line is probably due to the small expected frequencies within each group (for the most part less than one) since the EXAX2 comparisons in Section III revealed no beaker to beaker variation, except possibly in group 5.

The largest standardized value, corresponding to the beaker in group 5 with 100 percent mortality, lies somewhat above a straight line fitted to the other points. This suggests that it is larger than what would be expected of the extreme of the remaining values and thus may be too large to be due just to chance; i.e. it may be an outlier.

The extreme point corresponds to Beaker A of group 5. The observed mortality rate there is 100 percent, where the normal approximation is least reasonable. Thus before we infer that the point is in fact an outlier we should compare the observed mortality rate in this beaker with that in its companion beakers to determine if there is any statistical evidence of differences. One such comparison was made among the three beakers in group 5 as part of the EXAX2 analysis in Section III. There was a significant difference at the $\alpha = 0.07$ level. Such a significance level does not offer much statistical evidence of an outlier, especially when compared with what might be expected of the most extreme significance level of six independent tests, even when nothing was going on. We construct several more sensitive comparisons below.

Consider the results in group 5.

Group 5

	Replicate			
	A	B	C	
Dead	5	2	1	8
Live	0	3	4	7
	5	5	5	15

Beaker A is the suspected outlier. Compare its results to those from the other two beakers pooled.

	Replicate A	Replicate B,C	
Dead	5≡x	3	8
Live	0	7	7
	5	10	15

We can carry out an exact test of homogeneity of responses by means of the Fisher-Irwin test. (See Lehmann [5], Section 4.5, Lieberman and Owen [6].) If Beaker A in fact has the same mortality probability as B,C the probability of observing a table as extreme as the one above just due to chance can be calculated from the hypergeometric distribution as

$$P\ (X \geq 5) = \binom{8}{5} / \binom{15}{5} = 0.0186$$

Thus the approximate observed two tailed significance level is $2(0.0186) = 0.037$.

Now Beaker A was not chosen a priori. Taking selection into account we have, assuming homogeneous mortality probabilities,

P (most extreme of 3 beakers more significant than 0.037 level) =
P (at least one beaker more significant then 0.037 level) $\leq 3(0.037) = 0.11$

Thus there is perhaps marginal statistical evidence that the response rate in Beaker A of group 5 differs significantly from the response rates in the other beakers in that group.

Now group 5 was not chosen a priori. Taking selection of group 6 into account we have, assuming homogeneity among beakers within groups, for all groups,

P (most extreme of 6 groups more significant than 0.11 level) $\leq 6(0.11) = 0.66$

Thus when we account for selection of both group and beaker within group we must conclude that there is no statistical evidence of an outlying response.

A somewhat more powerful test, but one based on much more tenuous assumptions, can be constructed as follows. Based on the calculated response probabilities within each group and on the appearance of Figure II.8, suppose we assume that there is no trend in response probabilities in groups 1-5. (This is a very strong assumption.) Then we can pool the results across 14 beakers to compare with Beaker A of group 5. This results in the following table.

	5A	Fourteen other beakers in groups 1-5	
Dead	5≡x	10	15
Live	0	60	60
	5	75	75

We again carry out an exact test of homogeneity of responses by means of the Fisher-Irwin test. If Beaker 5A in fact has the same underlying mortality probability as those in each of the other 14 beakers in groups 1-5, the probability of observing a table as extreme as the one above just due to chance can be calculated from the hypergeometric distribution as

$$P\ (X \geq 5) = \frac{\binom{15}{5}}{\binom{75}{5}} = 1.74 \times 10^{-4}$$

Thus the approximate two tailed significance level is $2(1.74 \times 10^{-4}) = 3.48 \times 10^{-4}$. Now Beaker 5A was not chosen a priori. Taking selection of beakers into account we have, assuming homogeneity of response rates among all beakers in groups 1-5,

P (most extreme of 15 beakers more significant than 3.48×10^{-4} level)

$\leq 15\ (3.48 \times 10^{-4}) = 0.0052$

Thus, under the assumptions of this test, there is strong statistical evidence that the response in Beaker 5A is an outlier. However the assumption of constant mortality rates across all five groups is a bit too strong. In particular Beakers 5B,C show a 30 percent average mortality rate while the beakers in groups 1 to 4 have an average 12 percent mortality rate. While these rates are not statistically significantly different, they do at least suggest that the assumption of constant mortality rates across the first five groups is questionable.

In summary, the classification of the response in Beaker 5A as an outlier is equivocal. With the small sample sizes at hand the outlier tests are not significant (except under quite stringent assumptions). Thus in the absence of specific reasons to the contrary we will consider the responses in this beaker to be valid. Furthermore in the absence of overall beaker to beaker heterogeneity within groups we do not make any data adjustments. We carry out subsequent analyses with the data as presented.

C. OUTLIER DETECTION PROCEDURES APPLIED TO REPRODUCTION RESPONSES

Offspring are counted on a per beaker basis. Thus in the absence of a theoretical model upon which to base variance estimates (analagous to the binomial model for mortality data) or an assumption about the magnitude of beaker to beaker variation (e.g. that it is no greater than that for lengths)

there is no basis on which to augment the beaker to beaker variation estimates with information about within beaker variation (such as we did for mortality and length responses). Since such assumptions about theoretical models or about beaker to beaker variation are at best tenuous, we do not attempt them here and instead calculate variability estimates based on total results within each beaker (i.e. on a per beaker basis).

Outlier detection procedures are based on the residuals from one way analysis of variance models. Let R_{ij} denote the 21 day cumulative offspring per survivor in the j-th beaker of the i-th group. (In Chapman's test, with individual daphnids per beaker, attention is confined to those individual daphnids which survived for 21 days.) The one way analysis of variance model,

$$R_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1,\ldots, I \qquad j = 1,\ldots, J$$

was specified where R_{ij} denotes the total offspring, α_i is the fixed group effect, and ε_{ij} is the experimental variation. It is assumed that ε_{ij} are independent $N(0, \sigma_\varepsilon^2)$. The model was fitted to the data using the computer program BMDP1R in the BMDP statistical computing system [7]. The model fits and associated residual displays for the various data sets are presented below.

Figures VI.3 - VI.5 display the analysis of variance fit to the 21 day cumulative reproduction data in LeBlanc's Test A, the residuals plotted by group, and a normal probability plot of the ordered residuals. The individual responses are shown in Figure II.11. One of the residuals in group 5 stands out from the others in Figure VI.4. This residual does not lie on a straight line fitted by eye to the remaining residuals in Figure VI.5. It must thus be considered a potential outlier. To determine whether there is any statistical evidence that this extreme observation is in fact an outlier we can test whether the most extreme of 18 independent normally distributed random variables with mean 0 and standard deviation 20.64 is likely to exceed 49.25 in absolute value. (The extreme value is 189 and thus corresponds to a deviation of 49.25 from its group average. The four responses corresponding to group 7 have been deleted from the calculations below because there is virtually 100 percent mortality in group 7 and so reproduction patterns may not be comparable to those from other groups. The standard deviation estimate is thus $((9667.25 - (49.25)^2)/17)^{1/2} = 20.64$. We account for the correlations among the residuals from the same group by treating the $J \equiv 4$ correlated residuals as if they were $J-1 \equiv 3$ independent normal deviates). Thus

P [most extreme of 18 independent random deviates is greater than 49.25 in absolute value] $= 1 - [p\ (-49.25 < X < 49.25)]^{18} = 1 - \left[2\Phi\left(\frac{49.25}{20.64}\right) - 1\right]^{18}$

$$= 1 - [2\Phi\ (2.386) - 1]^{18} = 1 - (0.9830)^{18} = 0.27 \quad .$$

There is thus no statistical evidence that this extreme value is greater than what may be expected just due to random variation.

Figures VI.6 - VI.8 display the analysis of variance fit to the 21 day cumulative reproduction data in LeBlanc's Test B, the residuals plotted by group, and a normal probability plot of the ordered residuals. The individual responses are shown in Figure II.13. One of the residuals in group 3 stands out from the others in Figure VI.7. This residual does not lie on a straight line fitted by eye to the remaining residuals in the normal probability plot in Figure VI.8. It is thus a possible outlier. To determine whether there is any statistical evidence that this extreme observation is in fact an outlier we can test whether the most extreme of 7(4-1) = 21 independent normally distributed random variables with mean 0 and standard deviation 18.56 (i.e. $[(9697.5 - 53^2)/20]^{1/2}$) is likely to exceed 53 in absolute value. (The extreme value is 169 and thus corresponds to a deviation of 53 from its group average.) Thus

P [most extreme of 21 independent random normal deviates is greater than 53.0 in absolute value] $= 1 - [P(-53 < X < 53)]^{21} = 1 - [2\Phi(53/18.56) - 1]^{21}$

$= 1 - [2\Phi(2.856) - 1)^{21} = 1 - (0.9957)^{21} = 0.09$

This is at most marginally statistically significant and the situation is borderline. We will retain this observation in future analyses.

It is interesting to note the great similarity in reproduction results in Tests A and B. In each case the potential outlying value was on the high side. Perhaps the characteristics of these two daphnids and conditions in these beakers should be noted and repeated in future tests, to the extent possible, so as to obtain increased productivity.

Figures VI.9 - VI.11 display the analysis of variance fit to the 21 day cumulative reproduction data in Adams' selenium test, the residuals plotted by group, and a normal probability plot of the ordered residuals. The individual responses are shown in Figure II.15. Since groups 5-8 have 100 percent mortality we delete the residuals from these groups from the normal probability plot. None of the residuals in groups 1-5 stand out from the rest. There is thus no suggestion of outlying responses.

Figures VI.12 - VI.14 display the analysis of variance fit to the 21 day cumulative reproduction data in Chapman's beryllium test, the residuals plotted by group, and a normal probability plot of the ordered residuals. The individual responses are shown in Figure II.17. Since in this test the daphnids are housed individually, the reproduction figures pertain to individual daphnids. Comparisons are restricted to those daphnids that survived to the end of the test. None of the residuals stand out from the group. There is thus no suggestion of outlying responses.

Figures VI.15 - VI.17 display the analysis of variance fit to the 21 day cumulative reproduction data in Goulden's isophorone test, the residuals plotted by group, and a normal probability plot of the ordered residuals. The individual responses are shown in Figure II.22. Attention is confined to the individually housed daphnids that survived to the end of the test. Since just one daphnid in group 6 survived to the end of the test (and

produced no offspring), this response is deleted from the calculations below because the fertility patterns in this highest concentration group may not be comparable with those from the other groups. In groups 1-5, just one individually housed daphnid died. One of the residuals in group 1 stands out from the others in Figure VI.16. This residual does not lie on a straight line fitted by eye to the remaining residusls in Figure VI.17. It is thus a possible outlier. (This point corresponds to a daphnid that produced 135 offspring, while the average production in the control group was 77.29.) To determine whether there is any statistical evidence that this extreme point is in fact an outlier we can test whether the most extreme of 5(7-1) - 1 = 29 independent normally distributed random variables with mean 0 and standard deviation 14.38 (i.e. $[(9120.548 - 57.71^2)/28]^{1/2}$) is likely to exceed 57.71 in absolute value. (135 - 77.29 = 57.71). Thus

P [most extreme of 29 independent random deviates is greater than 57.71 in absolute value] $= 1 - [P\ (-57.71 < X < 47.71)]^{29} = 1 - [2\Phi\ (57.71/14.38) - 1]^{29}$

$$= 1 - [2\Phi\ (4.013) - 1]^{29} = 1 - (0.9999)^{29} = 0.002 \quad .$$

(Note that it might be argued that 14.38 is an underestimate of variability since it excludes the extreme residual, 57.71. If we repeat the above probability calculation using the standard deviation estimate 17.73 (from Figure VI.15), the extreme point is significant at the $\alpha = 0.03$ level). There is thus statistical evidence that this point is in excess of what is to be expected just due to random variation.

The above calculations, coupled with the appearances of Figures II.22, VI.16 and VI.17 suggest that this point be deleted from subsequent comparisons. This decision may well impact on whether groups 4 and 5 are considered to differ significantly from the control group. As such, the decision as to whether to include or exclude this point from subsequent comparisons should also be based on biological judgement and knowledge of experimental details. If the extreme result represents normal biological variation then perhaps the response should be considered with the others. Outlier detection procedures are merely screening devices to direct attention to those places where biological judgement should be applied.

For the sake of illustration we have chosen to exclude this observation from subsequent comparisons. In actual statistical analyses to support regulatory applications or regulatory decisions the subsequent analyses might be carried out both with this point in and out. Any biologically important differences in analysis results would need to be resolved on biological grounds.

D. OUTLIER DETECTION PROCEDURES APPLIED TO LENGTH RESPONSES

Lengths are measured at periodic intervals on individuals daphnids. In the data sets discussed in this report, length determinations were made by LeBlanc (in Tests A and B) on days 7 and 21 and by Chapman (in his test on beryllium) on day 21. In LeBlanc's tests we can think of outlying beaker averages relative to their group averages as well as outlying individual lengths relative to their beaker averages. In Chapman's test each beaker contains just a single daphnid and so we need only consider outlying individual lengths around their group means.

We first consider outlier detection procedures for the beaker averages about their group means in LeBlanc's Tests A and B. From the discussion in Subsection V.D we associate with the beaker averages estimates of variability based on the mean squares for beakers within group 5, with degrees of freedom $I(J\underset{\sim}{x}-1)$ as discussed there. Let N_j denote the sample size within the j-th beaker of i-th group, $N = \Sigma_j N_j$, and let σ_b^2, σ_c^2 denote the components of variance due to beakers and daphnids with beakers respectively. Let $\bar{X}_j$ denote the average length within the j-th beaker and $\bar{\bar{X}} \equiv \Sigma_j N_j \bar{X}_j/N$ denote the (weighted) average daphnid length within the i-th group. The standard error of $\bar{X}_j$ is $[(\sigma_c^2 + N_j \sigma_b^2)/N_j]^{1/2}$. We approximate $\sigma_c^2 + N_j \sigma_b^2$ by the mean square for beakers within groups, MSI, as discussed in Subsection V.D. (This approximation would be exact if all the sample sizes, N_j, were equal across beakers and across groups.) The values of MSI were calculated in Subsection V.D to be

Test A:	MSI = 0.2722	with 98 degrees of freedom
Test B:	MSI = 0.2707	with 141 degrees of freedom

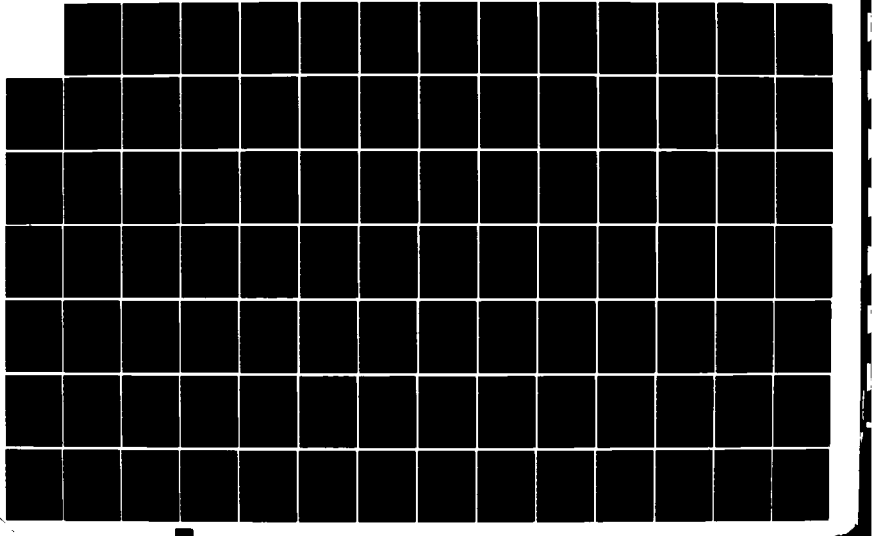
The residuals of the beaker averages about their group averages can be calculated from Figures IV.2 - IV.7 for Test A and from Figures IV.9 - IV.15 for Test B. The standard errors of the residuals are approximated by $[(1 - N_j/N)\ MSI/N_j]^{1/2}$. (This approximation would be exact if all the N_j's were equal across beakers and across groups.) Tables VI.1 and VI.2 contain the ordered residuals multipled by the factors $N_j^{1/2}\ (1 - N_j/N)^{-1/2}$, so as to have approximately constant variance. These tables also contain group and beaker identification and plotting positions appropriate for preparing normal probability plots. The normal probability plots of these standardized residuals appear in Figures VI.18 and VI.19. The reference lines on these plots correspond to normal distributions with mean 0 and with variance MSI. In both tests the beaker averages conform nicely to the reference lines. Thus there is no suggestion of outlying average beaker lengths within any treatment or control group.

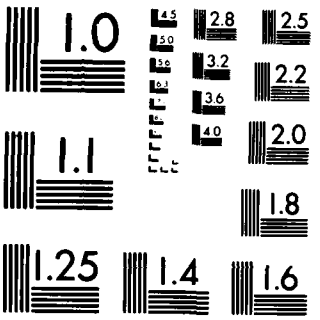
We next consider outlier detection procedures for individual (21 day) beaker lengths. In LeBlanc's tests this involves studying the deviations of individual lengths within each beaker about their beaker averages. In Chapman's test this involves studying the deviations of individual daphnid lengths about their group averages.

The approach is much like that used for outlier detection with reproduction responses. Namely the outlier detection procedures are based on

AD-A152 070 DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA(U) BATTELLE COLUMBUS LABS OH P I FEDER DEC 81 BATT-G-7673 DAMD17-80-C-0165 3/5

UNCLASSIFIED F/G 6/20 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

TABLE VI.1. LE BLANC TEST A - ORDERED STANDARDIZED RESIDUALS OF BEAKER AVERAGE LENGTHS FROM OVERALL GROUP AVERAGE LENGTHS-TO DETECT OUTYLYING BEAKER AVERAGES (GROUPS 1-6 ONLY)

Order	Group	Beaker	Standardized Residual	Plot Position	Order	Group	Beaker	Residual	Plot Position
1	3	C	-1.05	2.1	13	5	B	-0.03	52.1
2	1	A	-0.80	6.2	14	2	D	0.12	56.2
3	4	C	-0.68	10.4	15	6	D	0.20	60.4
4	5	C	-0.47	14.6	16	3	D	0.26	64.6
5	1	D	-0.46	18.7	17	4	A	0.29	68.7
6	3	B	-0.41	22.9	18	6	C	0.33	72.9
7	5	D	-0.36	27.1	19	2	A	0.36	77.1
8	2	B	-0.24	31.2	20	4	D	0.44	81.2
9	2	C	-0.24	35.4	21	1	B	0.63	85.4
10	6	B	-0.23	39.6	22	1	C	0.63	89.6
11	4	B	-0.07	43.7	23	5	A	0.85	93.7
12	6	A	-0.06	47.9	24	3	A	1.11	97.9

TABLE VI.2. LE BLANC TEST B - ORDERED STANDARDIZED RESIDUALS OF BEAKER AVERAGE LENGTHS FROM OVERALL GROUP AVERAGE LENGTHS-TO DETECT OUTLYING BEAKER AVERAGES

Order	Group	Beaker	Standardized Residual	Plot Position	Order	Group	Beaker	Standardized Residual	Plot Position
1	6	B	-1.07	1.8	15	7	B	-0.06	51.8
2	3	C	-0.99	5.4	16	4	C	-0.03	55.4
3	2	C	-0.80	8.9	17	5	A	0.02	58.9
4	4	D	-0.55	12.5	18	1	D	0.03	62.5
5	3	A	-0.39	16.1	19	4	A	0.13	66.1
6	2	B	-0.34	19.6	20	7	C	0.36	69.6
7	7	D	-0.33	23.2	21	1	A	0.37	73.2
8	7	A	-0.29	36.8	22	5	D	0.39	76.8
9	5	C	-0.27	30.4	23	4	B	0.44	80.4
10	1	B	-0.20	33.9	24	6	A	0.49	83.9
11	1	C	-0.20	37.5	25	3	B	0.63	87.5
12	5	B	-0.14	41.1	26	6	D	0.74	91.1
13	6	C	-0.12	44.6	27	3	D	0.79	94.6
14	2	A	-0.11	48.2	28	2	D	1.20	98.2

the residuals from one way analysis of variance models. For LeBlanc's Tests A and B consider each beaker (with one or more survivors at 21 days) as a separate classification, irrespective of group. Thus for Test A there are 25 classifications while for Test B there are 28 classifications. Let L_{jk} denote the 21 day length of the k-th surviving daphnid in the j-th beaker. For Chapman's test, with just one daphnid per beaker, let L_{jk} denote the 21 day length of the k-th surviving daphnid in the j-th (treatment or control group. The one way analysis of variance model

$$L_{jk} = \mu + \alpha_j + \varepsilon_{jk} \qquad j = 1,\ldots, J \qquad k = 1,\ldots, K_j$$

was specified where L_{jk} denotes the 21 day length, α_j is the fixed beaker or group effect, and ε_{jk} is the experimental random variation. It is assumed that the ε_{jk} are independent $N(0, \sigma_\varepsilon^2)$. The model was fitted to the data using the computer program BMDPIR in the BMDP statistical computing system [7]. The model fits and associated residual displays are presented below.

Figures VI.20 - VI.22 display the analysis of variance fit to the 21 day lengths of survivors in LeBlanc's Test A, the residuals plotted by classification, and a normal probability plot of the ordered residuals. None of the residuals appear to stand apart from the others in Figure VI.21. A straight line fitted by eye accomodates all the residuals in Figure VI.22. The residuals thus appear to follow an approximate normal distribution and there is no suggestion of any outlying lengths.

Figure VI.23 - VI.25 display the analysis of variance fit to the 21 day lengths of survivors in LeBlanc's Test B, the residuals plotted by classification, and a normal probability plot of the ordered residuals. Again, none of the residuals stand apart from the others in Figure VI.24 nor deviate from the straight line in Figure VI.25. Thus there is no suggestion of any outlying lengths in Test B.

Figures VI.26 - VI.28 display the analysis of variance fit to the 21 day lengths of survivors in Chapman's beryllium test, the residuals plotted by group, and a normal probability plot of the ordered residuals. Four of the residuals (one from group 5, one from group 6, and two from group 7) appear to be removed from the others, on the low side, in Figure VI.27. These residuals are also removed from a straight line fitted to the remaining ones in Figure VI.28. There is thus a definite suggestion that the four daphnids corresponding to these cases may be outliers on the low side. That is, those daphnids may be dwarfed, relative to their group averages, more than could be expected of the most extreme of 63 deviations just due to chance. To determine whether there is any statistical evidence that the most extreme of these residuals is in fact an outlier we can test whether the most extreme of 63 - 8 = 55 independently normally distributed random variables with mean 0 and standard deviation 0.275 (as determined from the straight line drawn to the majority of the values in FIgure VI.28) is likely to exceed 0.8167 in absolute value. (The extreme length is 3.0000 mm and the group average is 3.8167 mm.) Thus

P [most extreme of 55 independent random normal deviates is greater than 0.8167 in absolute value] $= 1 - [P(-0.8167 < X < 0.8167)]^{55}$

$= 1 - [2\Phi(0.8167/0.275) - 1)^{55}$

$= 1 - [2\Phi(2.9698) - 1]^{55} = 1 - (0.9970)^{55}$

$= 1 - 0.85 = 0.15$

The degree of statistical evidence is marginal and so the situation is borderline. Let's consider the second most extreme residual. This residual has value -0.7900.

P [second most extreme of 55 independent random normal deviates is greater than 0.79 in absolute value]

$= 1 - [P(-0.79 < X < 0.79)]^{55}$

$- 55 [P(-0.79 < X < 0.79]^{54} [2P(X > 0.79)]$

$= 1 - [2\Phi(0.79/0.275) - 1]^{55}$

$- 55 [2\Phi(0.79/0.275) - 1]^{54} [2(1 - \Phi(0.79/0.275))] = 0.02$

Thus the second extreme residual is more highly statistically significant. This suggests that there may be some physical reason underlying the extreme values on the low side. Basic records should be checked by the investigator to try to identify a reason. The proper action would depend on the physical explanation. In the absence of such information we choose to regard these observations as valid and carry out subsequent analyses incorporating them.

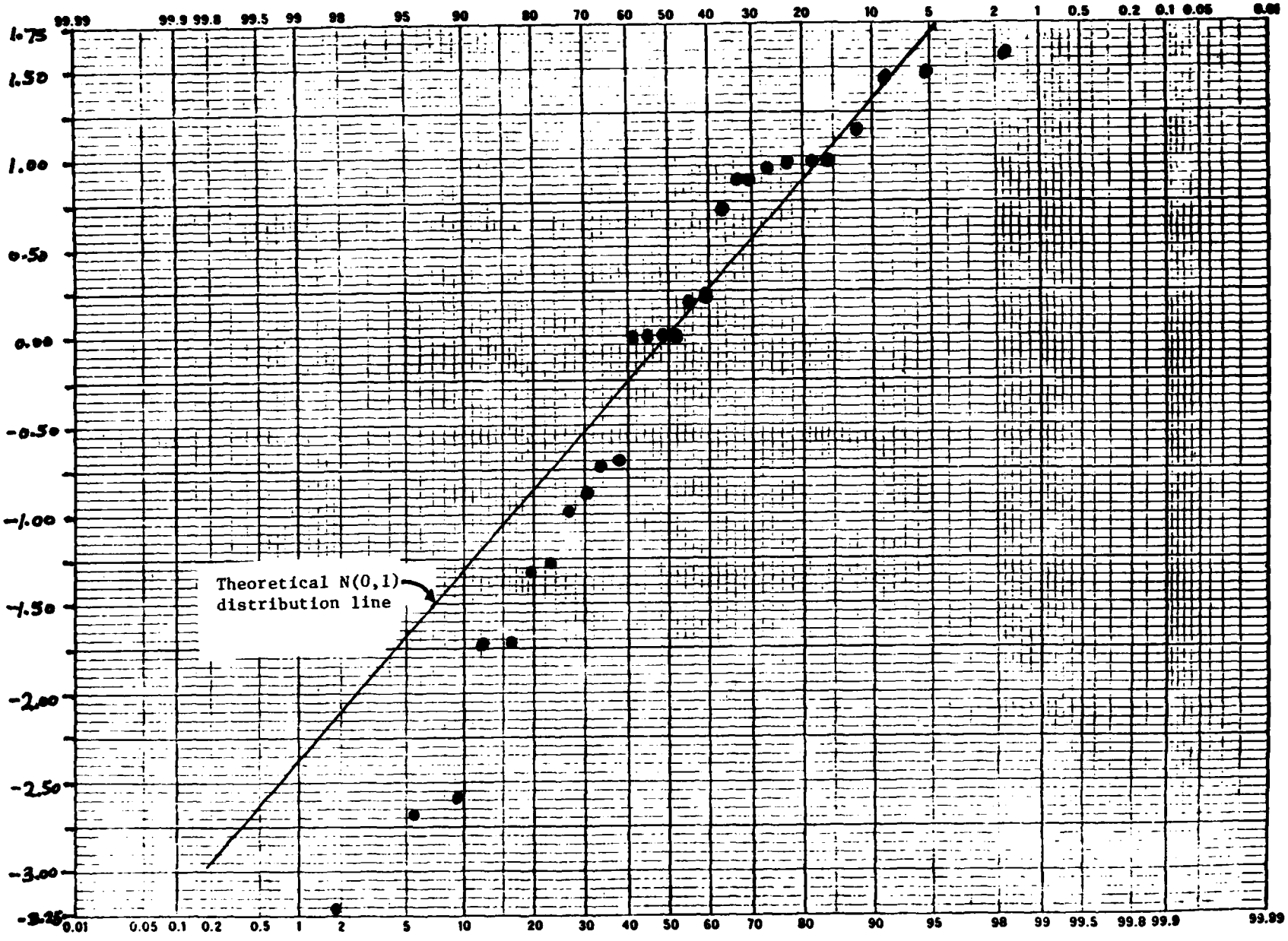


Figure VI.1. Normal probability plot of standardized values based on LeBlanc Test A - 21 day mortality--to detect outlying responses

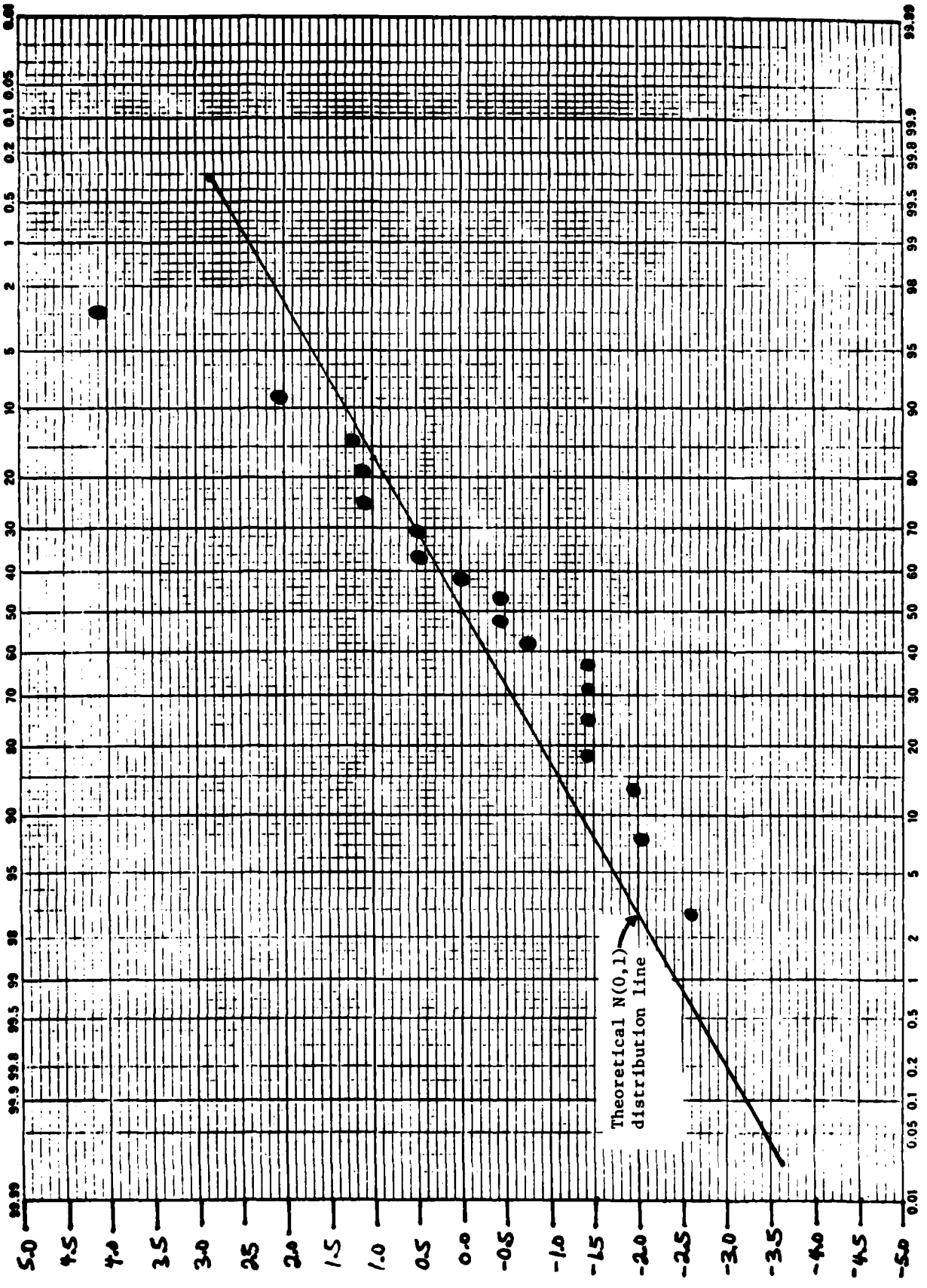


Figure VI.2. Normal probability plot of standardized values based on Goulden Isophorone test - 21 day mortality--to detect outlying responses

REGRESSION TITLE. . . . . .BIONOMICS TEST A, PROB. PLOTS FOR REPRODUCTION
DEPENDENT VARIABLE. . . . . 2 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9878 STD. ERROR OF EST. 21.4557
MULTIPLE R-SQUARE .9757

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	388371.750	7	55481.679	120.522	0.00000
RESIDUAL	9667.250	21	460.345		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1	3	102.50000	10.728	.325	9.555	-.000	1.000000
GRP2	4	154.00000	10.728	.488	14.355	0.000	1.000000
GRP3	5	128.25000	10.728	.407	11.955	0.000	1.000000
GRP4	6	129.00000	10.728	.409	12.025	0.000	1.000000
GRP5	7	139.75000	10.728	.443	13.027	0.000	1.000000
GRP6	8	101.25000	10.728	.321	9.438	-.000	1.000000
GRP7	9	.00000	10.728	.000	.000	1.000	1.000000

Figure VI.3. Analysis of variance fit of 21 day cumulative reproduction on group. LeBlanc test A.

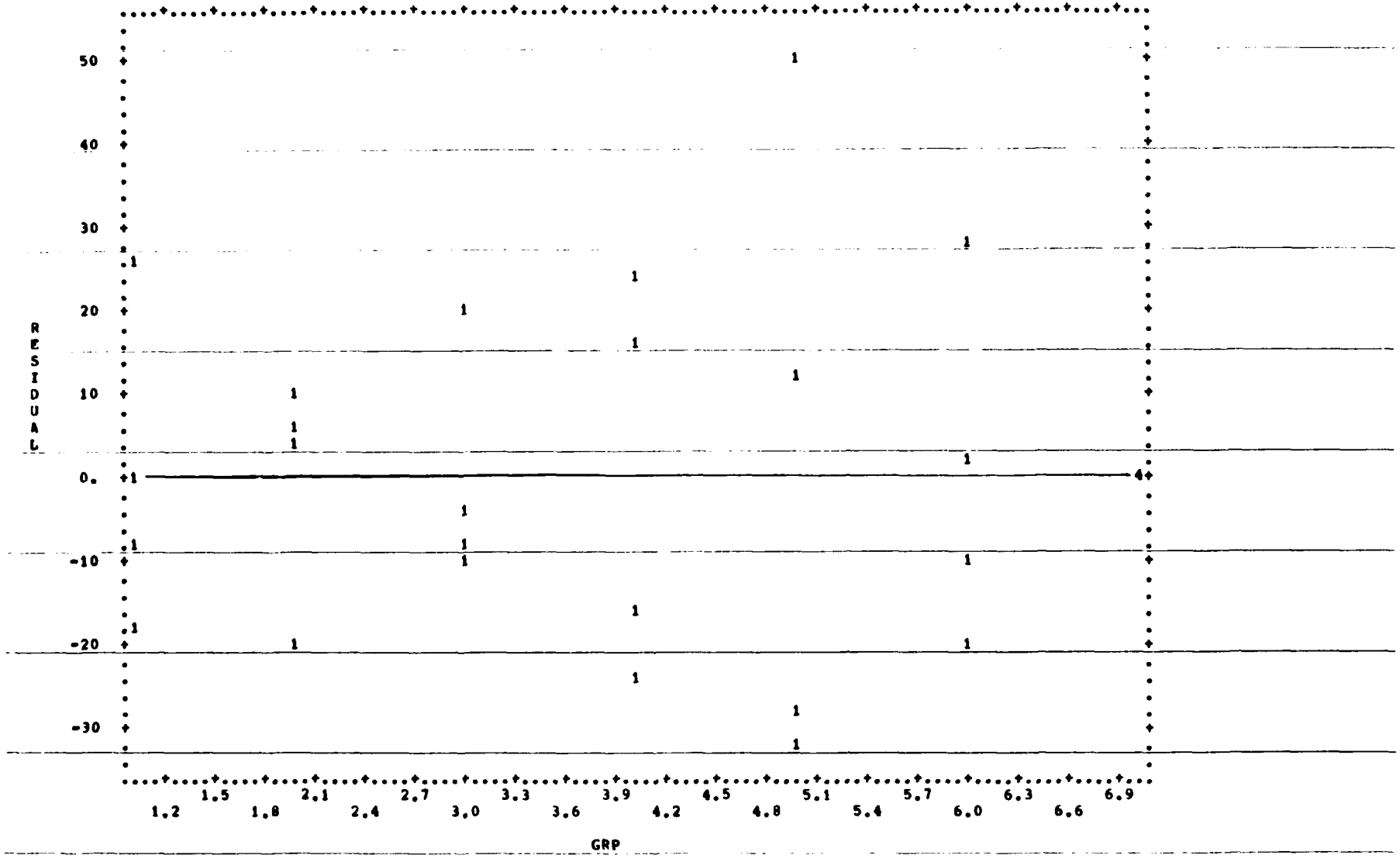


Figure VI.4. Residuals from analysis of variance fit versus group. LeBlanc test A.

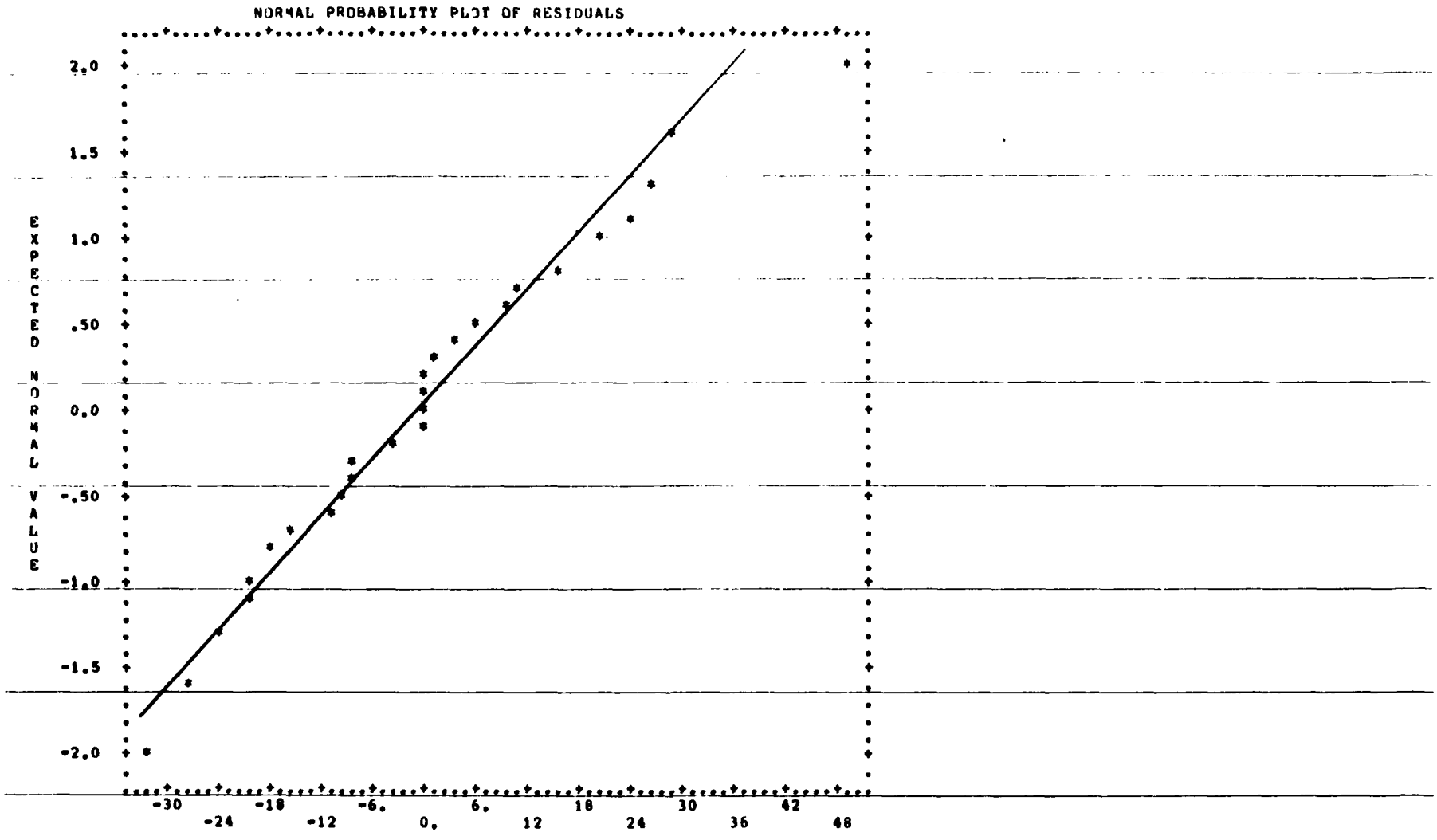


Figure VI.5. Normal probability plot of residuals from analysis of variance fit. LeBlanc test A.

REGRESSION TITLE. . . . . .BIONOMICS TEST B, PROB. PLOTS FOR REPRODUCTION
DEPENDENT VARIABLE. . . . . 2 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9878 STD. ERROR OF EST. 21.4892
MULTIPLE R-SQUARE .9758

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	391040.500	7	55862.929	120.972	0.00000
RESIDUAL	9697.500	21	461.786		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1	3	111.50000	10.745	.352	10.377	0.000	1.000000
GRP2	4	107.00000	10.745	.338	9.958	-.000	1.000000
GRP3	5	116.00000	10.745	.366	10.796	0.000	1.000000
GRP4	6	134.50000	10.745	.425	12.518	0.000	1.000000
GRP5	7	128.25000	10.745	.405	11.936	0.000	1.000000
GRP6	8	119.25000	10.745	.377	11.099	0.000	1.000000
GRP7	9	108.00000	10.745	.341	10.052	0.000	1.000000

Figure VI.6. Analysis of variance fit of 21 day cumulative reproduction on group. LeBlanc test B.

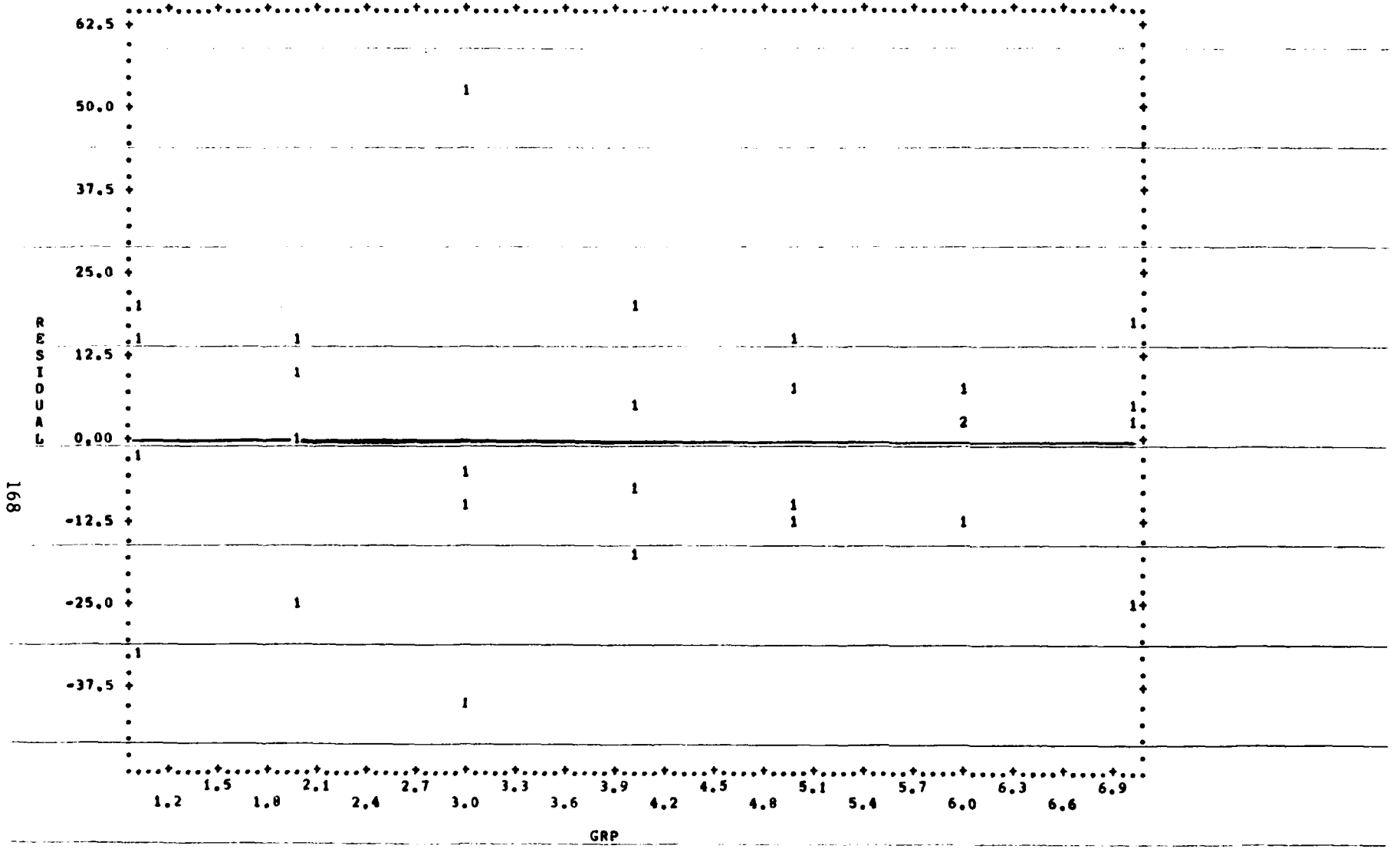


Figure VI.7. Residuals from analysis of variance fit versus group. LeBlanc test B.

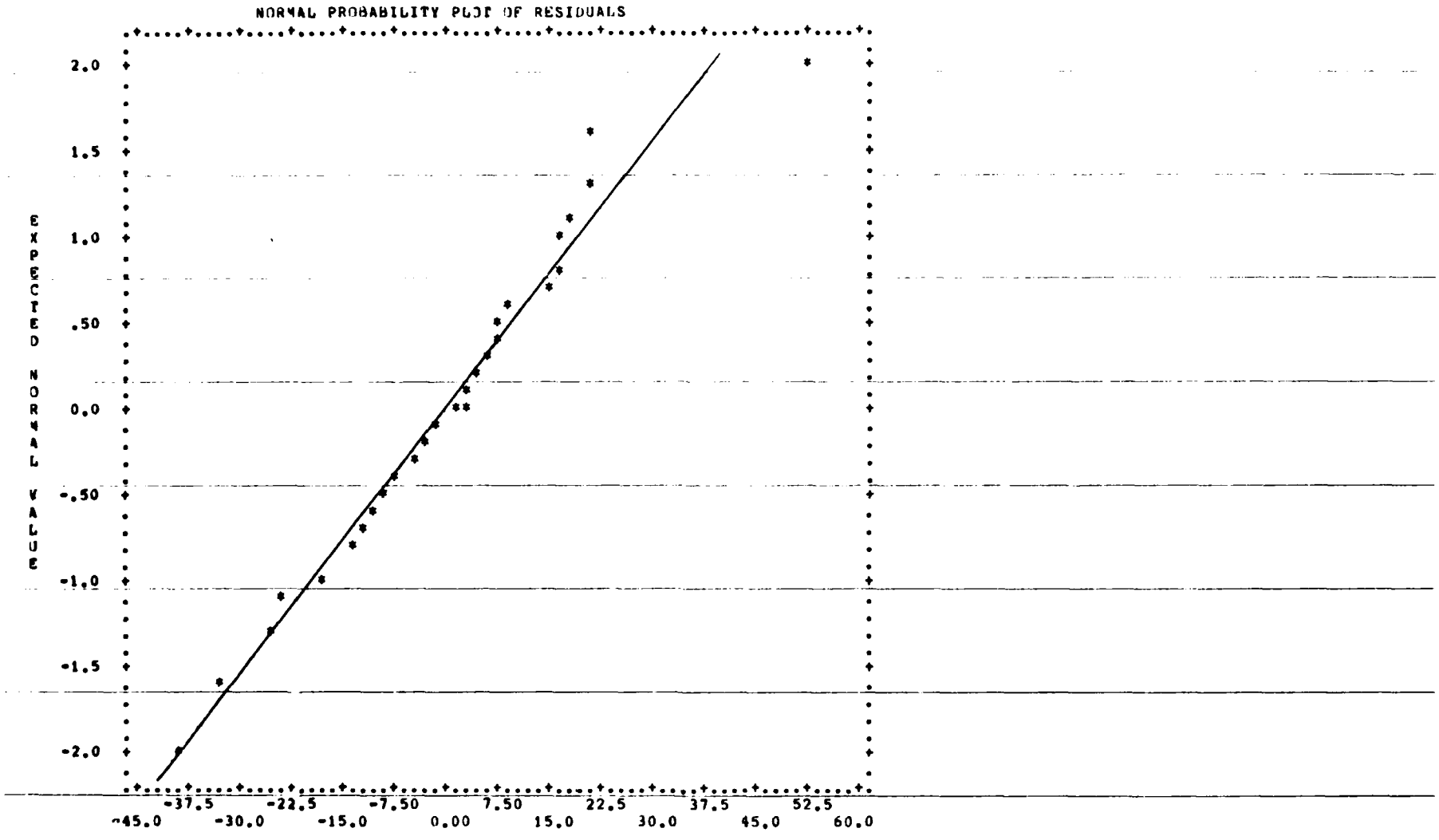


Figure VI.8. Normal probability plot of residuals from analysis of variance fit. LeBlanc test B.

REGRESSION TITLE. . . . . .SELENIUM TEST, PROB. PLOTS FOR REPRODUCTION
DEPENDENT VARIABLE. . . . . 2 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9832 STD. ERROR OF EST. 7.0151
MULTIPLE R-SQUARE .9666

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	22783.070	8	2847.884	57.871	.00000
RESIDUAL	787.380	16	49.211		

		VARIABLE	COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1		3	47.73333	4.050	.539	11.786	0.000	1.000000
GRP2		4	45.06667	4.050	.508	11.127	0.000	1.000000
GRP3		5	41.43333	4.050	.467	10.230	0.000	1.000000
GRP4		6	39.60000	4.050	.447	9.777	.000	1.000000
GRP5	*	7	.00000	4.050	.000	.000	1.000	1.000000
GRP6	*	8	.00000	4.050	.000	.000	1.000	1.000000
GRP7	*	9	.00000	4.050	.000	.000	1.000	1.000000
GRP8	*	10	.00000	4.050	.000	.000	1.000	1.000000

* 100% mortality in these groups

Figure VI.9. Analysis of variance fit of 21 day cumulative reproduction on group. Adams' selenium test.

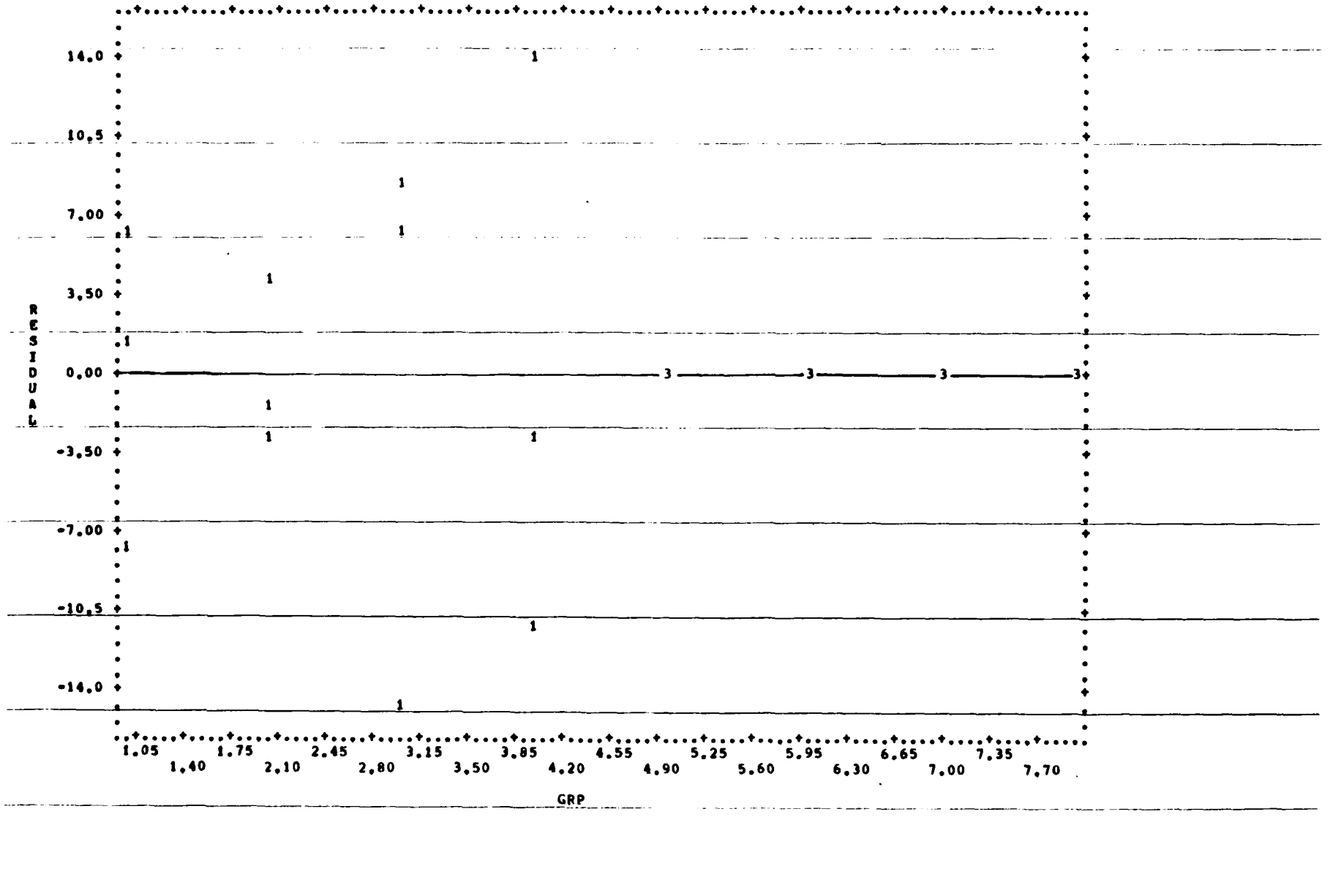


Figure VI.10. Residuals from analysis of variance fit versus group. Adams' selenium test.

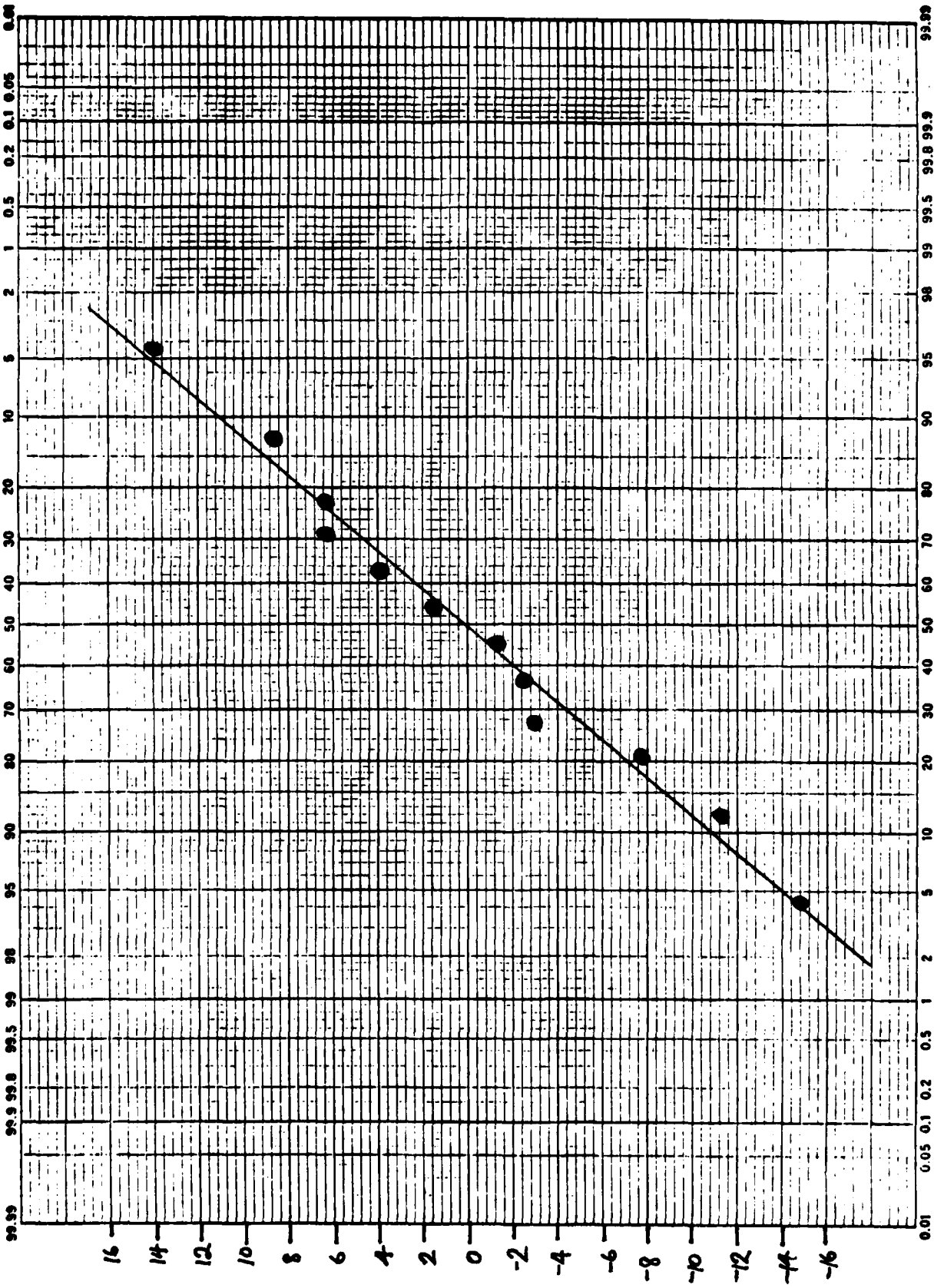


Figure VI.11. Normal probability plot of residuals from analysis of variance fit. Adams selenium test. Groups 1-4 only.

REGRESSION TITLE. . . . . .CHAPMAN TEST--PROB. PLOTS FOR REPRODUCTION
DEPENDENT VARIABLE. . . . . 13 NYOUNG
TOLERANCE 0100

ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9328 STD. ERROR OF EST. 48.5391
MULTIPLE R-SQUARE .8702

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	872238.267	8	109029.783	46.087	0.00000
RESIDUAL	130116.733	55	2365.759		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1	14	150.50000	15.381	.475	9.785	.000	1.000000
GRP2	15	177.50000	17.197	.501	10.322	0.000	1.000000
GRP3	16	88.25000	17.197	.249	5.132	.000	1.000000
GRP4	17	74.50000	19.857	.182	3.752	.000	1.000000
GRP5	18	95.66667	19.857	.234	4.818	.000	1.000000
GRP6	19	112.00000	16.213	.336	6.908	.000	1.000000
GRP7	20	92.40000	15.381	.292	6.007	.000	1.000000
GRP8	21	86.50000	19.857	.212	4.356	.000	1.000000

Figure VI.12. Analysis of variance fit of 21 day cumulative reproduction on group. Chapman beryllium test. Surviving daphnids only.

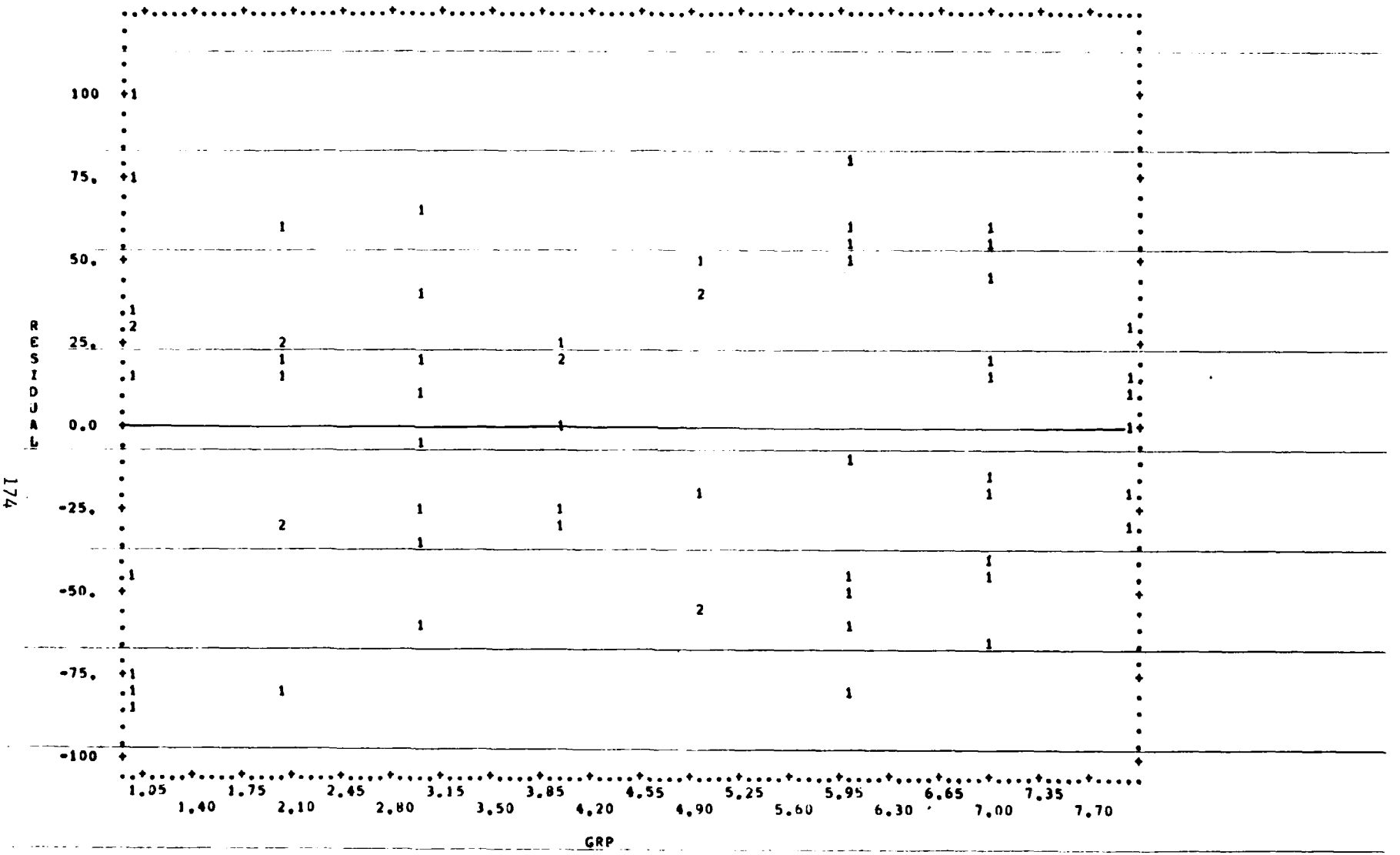


Figure VI.13. Residuals from analysis of variance fit versus group. Chapman beryllium test. Surviving daphnids only.

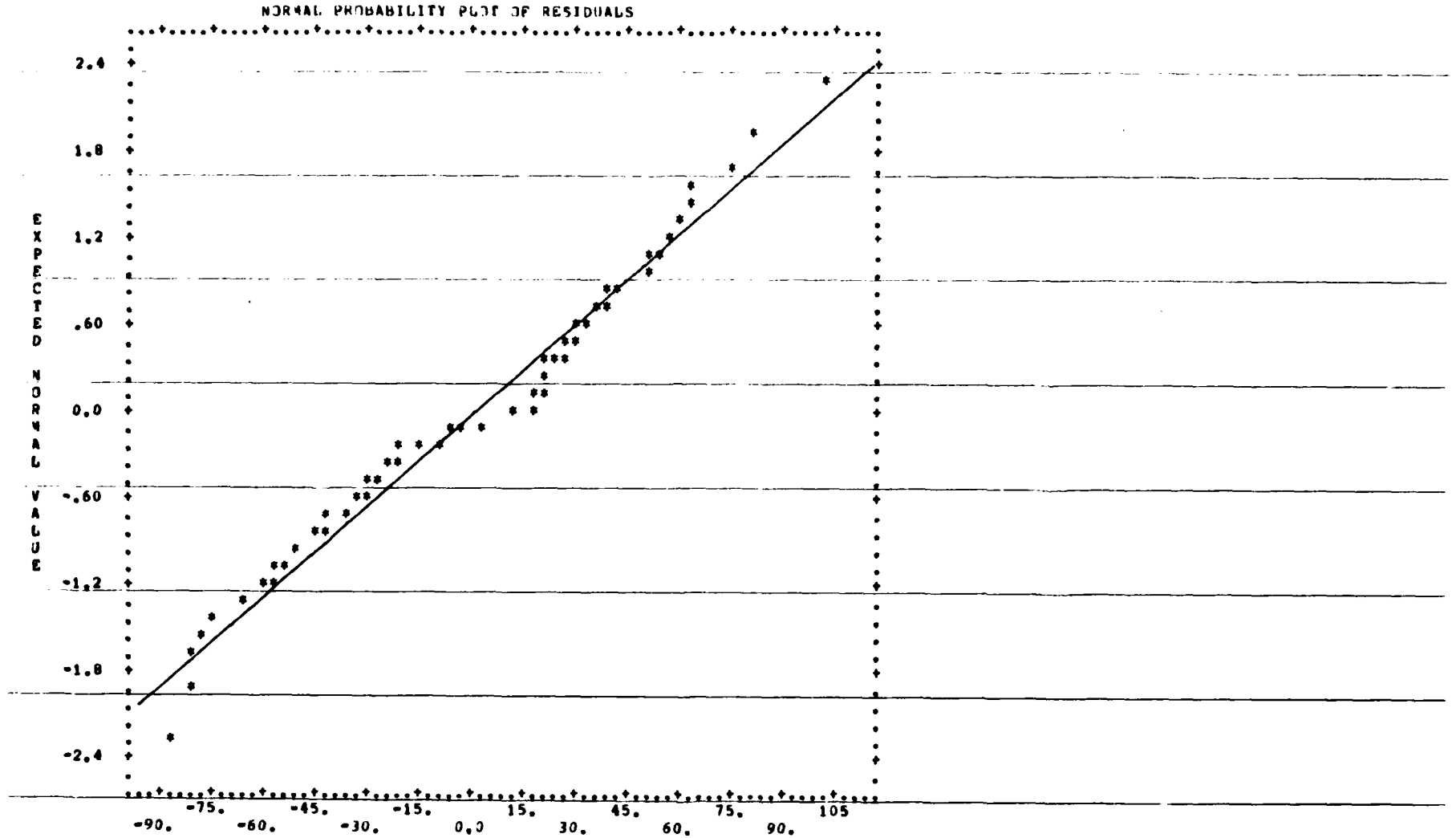


Figure VI.14. Normal probability plot of residuals from analysis of variance fit. Chapman beryllium test. Surviving daphnids only.

REGRESSION TITLE. . . . . .GOULDEN TEST--PROB. PLOTS FOR REPRODUCTION
DEPENDENT VARIABLE. . . . . 12 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9755 STD. ERROR OF EST. 17.7342
MULTIPLE R-SQUARE .9516

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	179399.452	6	29899.909	95.071	0.00000
RESIDUAL	9120.548	29	314.502		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1	13	77.28571	6.703	.471	11.530	0.000	1.000000
GRP2	14	86.83333	7.240	.490	11.994	0.000	1.000000
GRP3	15	90.85714	6.703	.554	13.555	0.000	1.000000
GRP4	16	69.00000	6.703	.420	10.294	0.000	1.000000
GRP5	17	13.28571	6.703	.081	1.982	.057	1.000000
GRP6	18	-.00000	17.734	-.000	-.000	1.000	1.000000

Figure VI.15. Analysis of variance fit of 21 day cumulative reproduction on group. Goulden isophorone test. Surviving, individually housed daphnids only.

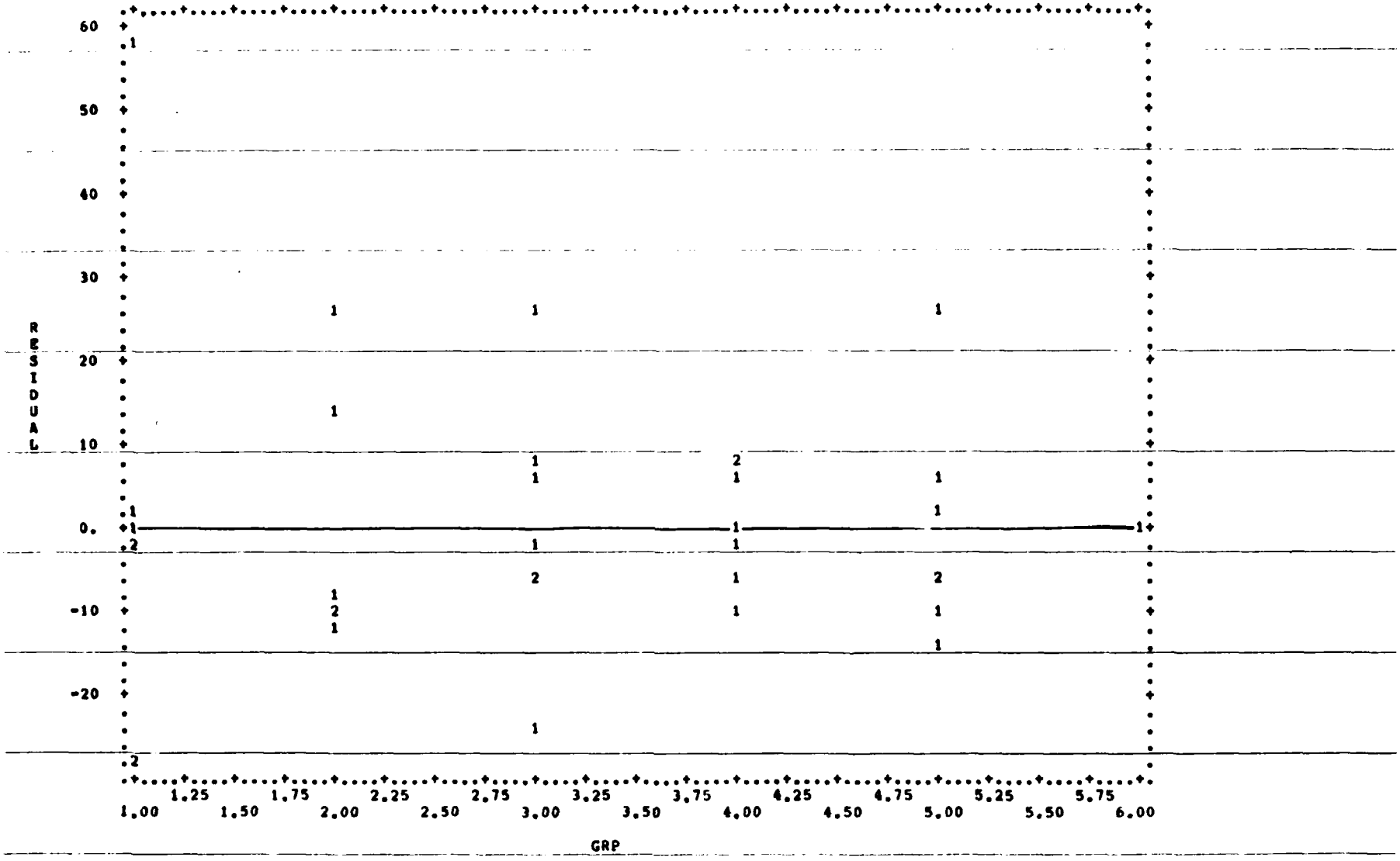


Figure VI.16. Residuals from analysis of variance fit versus group. Goulden isophorone test. Surviving, individually housed daphnids only.

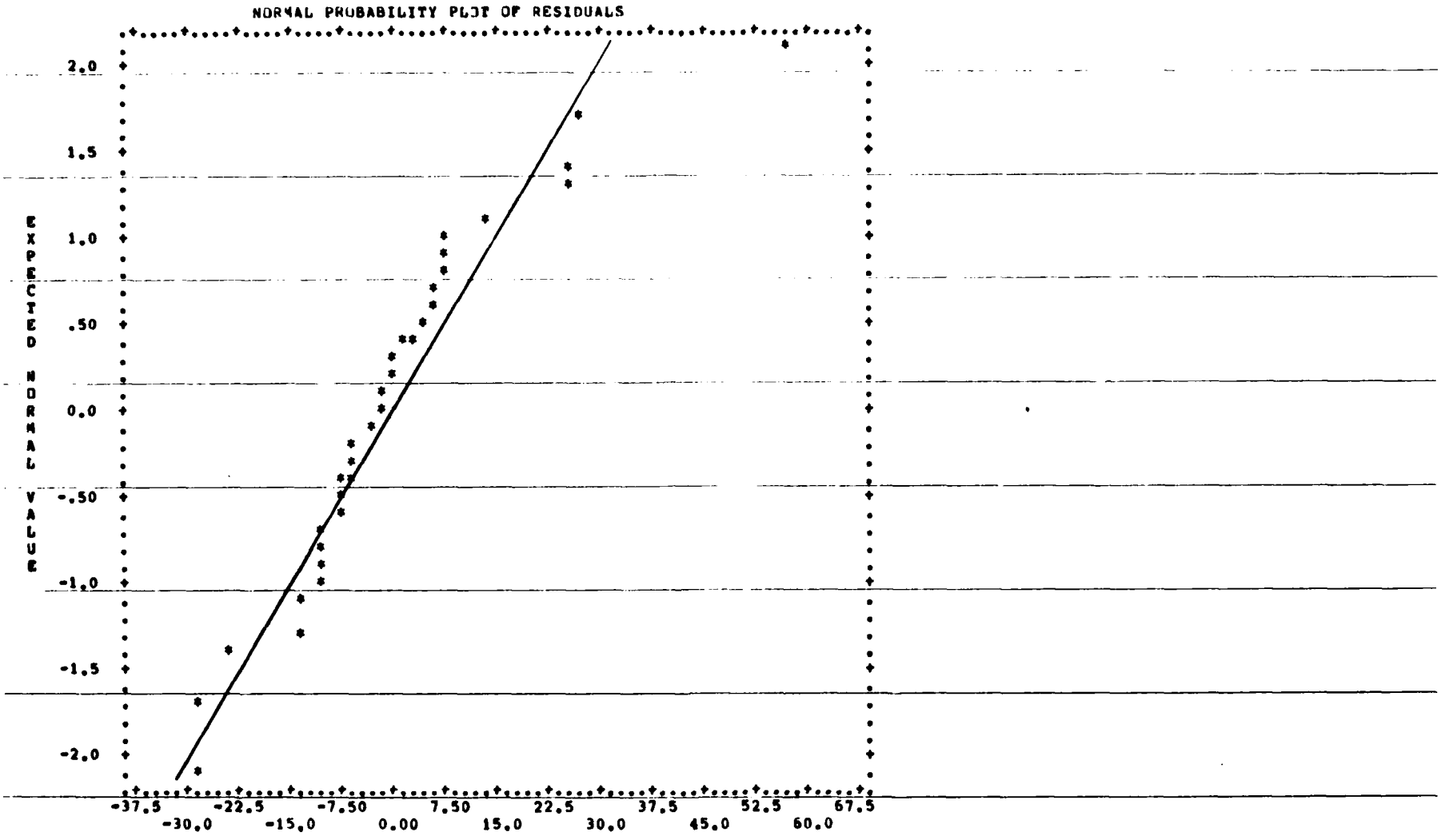


Figure VI.17. Normal probability plot of residuals from analysis of variance fit. Goulden isophorone test. Surviving, individually housed daphnids only.

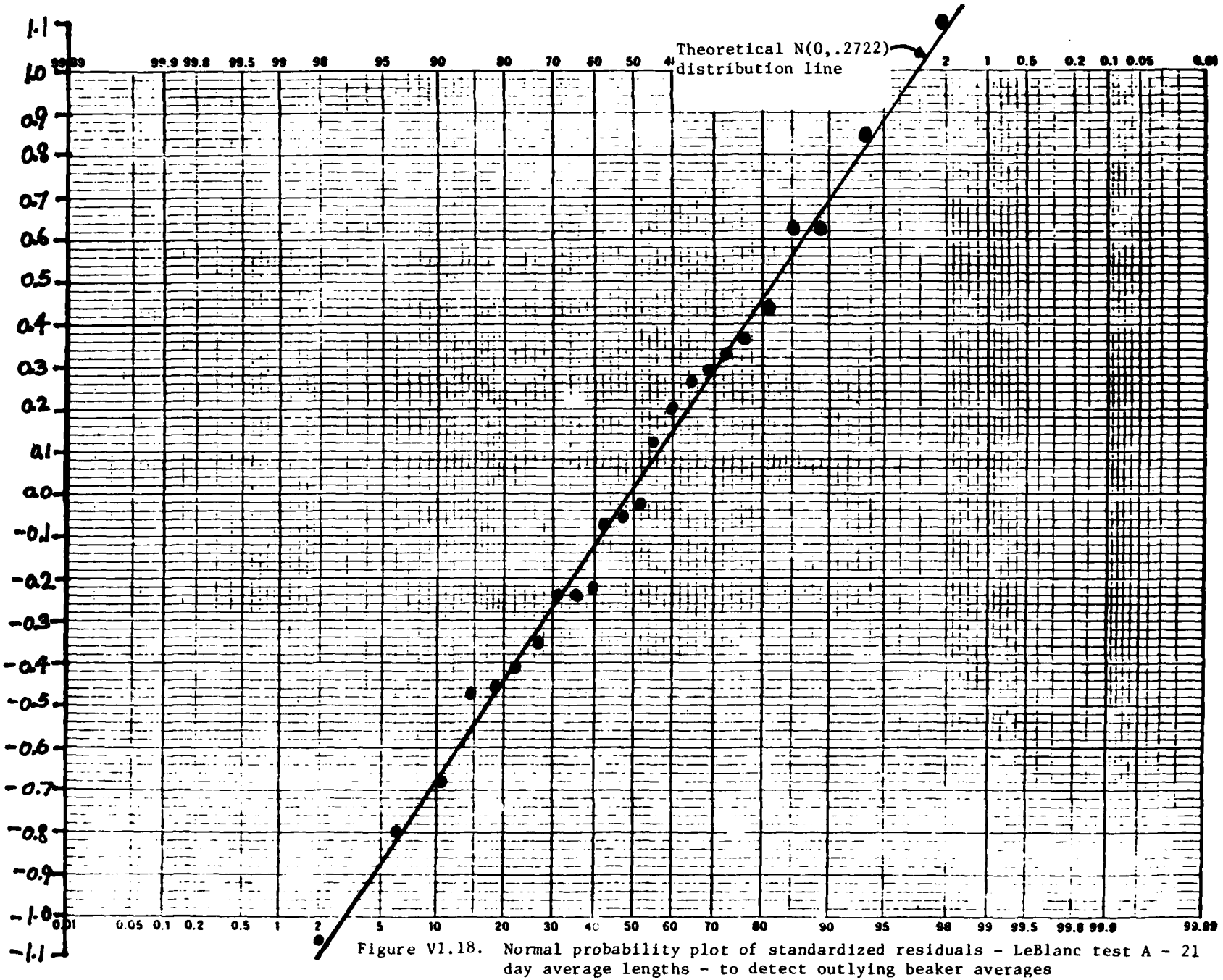


Figure VI.18. Normal probability plot of standardized residuals - LeBlanc test A - 21 day average lengths - to detect outlying beaker averages

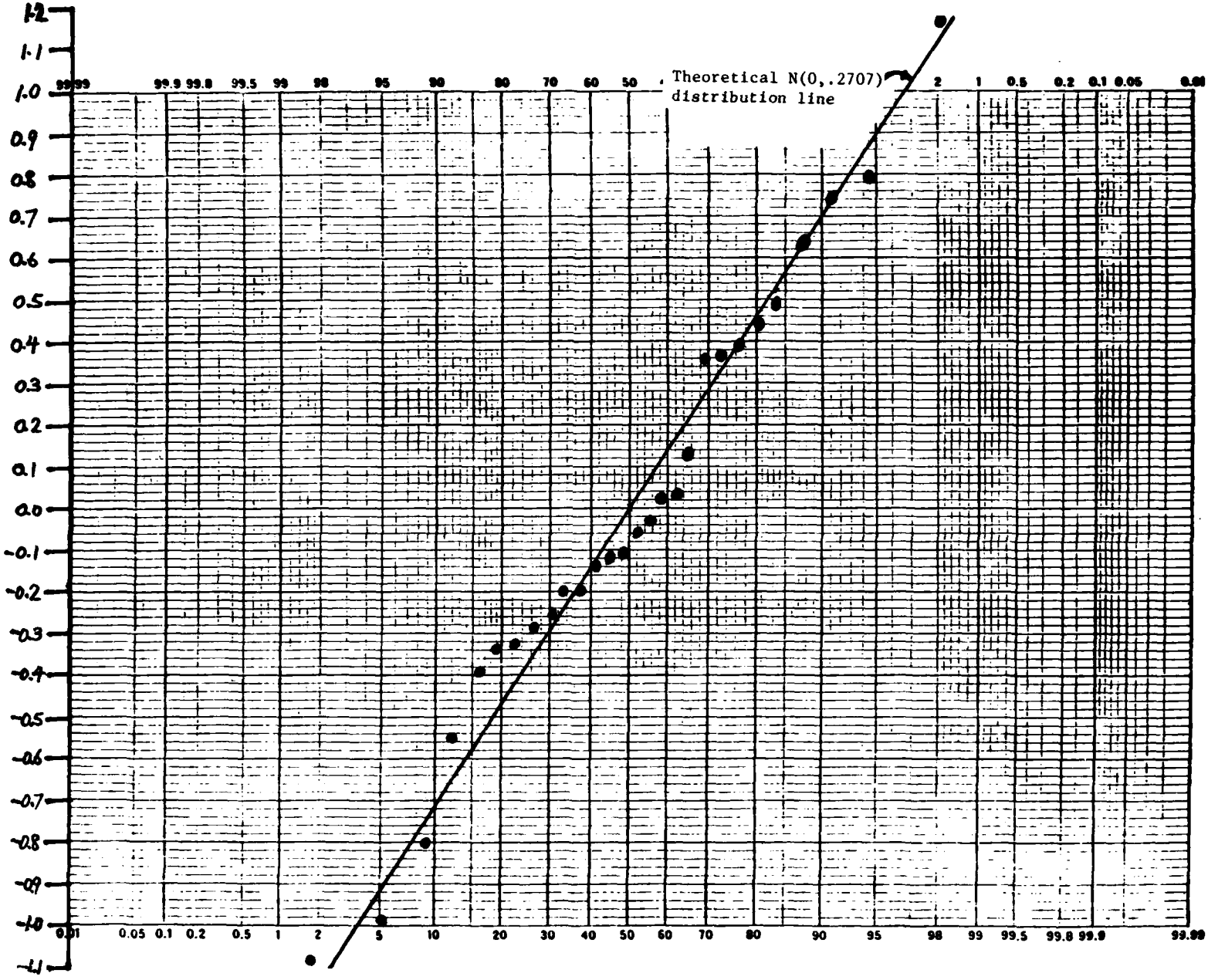


Figure VI.19. Normal probability plot of standardized residuals - LeBlanc Test B-21 day average lengths - to detect outlying beaker averages

REGRESSION TITLE. . . . . .BIONOMICS TEST A LENGTHS--PROB. PLOTS FOR TANK MEANS
DEPENDENT VARIABLE. . . . . 2 LENGTH
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SLE WP.TROP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9980 STD. ERROR OF EST. .3343
MULTIPLE R-SQUARE .9959

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	10280.800	25	411.232	3679.926	0.00000
RESIDUAL	42.130	377	.112		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
G1	3	4.80000	.081	.195	59.203	0.000	1.000000
G2	4	5.10000	.081	.207	62.963	0.000	1.000000
G3	5	5.10000	.081	.207	62.903	0.000	1.000000
G4	6	4.87059	.081	.198	60.073	0.000	1.000000
G5	7	5.35000	.079	.223	67.899	0.000	1.000000
G6	8	5.23000	.075	.230	69.967	0.000	1.000000
G7	9	5.23000	.075	.230	69.967	0.000	1.000000
G8	10	5.30000	.077	.227	69.108	0.000	1.000000
G9	11	5.18500	.075	.228	69.365	0.000	1.000000
G10	12	4.89474	.077	.210	63.824	0.000	1.000000
G11	13	4.73333	.086	.180	54.839	0.000	1.000000
G12	14	5.03125	.084	.198	60.202	0.000	1.000000
G13	15	5.16471	.081	.210	63.701	0.000	1.000000
G14	16	5.09000	.075	.224	68.094	0.000	1.000000
G15	17	4.95882	.081	.201	61.162	0.000	1.000000
G16	18	5.16947	.077	.223	67.667	0.000	1.000000
G17	19	5.13889	.079	.215	65.220	0.000	1.000000
G18	20	4.96000	.075	.218	66.355	0.000	1.000000
G19	21	4.86250	.084	.191	58.183	0.000	1.000000
G20	22	4.89474	.077	.210	63.824	0.000	1.000000
G21	23	4.98889	.079	.208	63.316	0.000	1.000000
G22	24	4.95625	.084	.135	59.305	0.000	1.000000
G23	25	5.14000	.149	.113	34.381	0.000	1.000000
G24	26	5.20000	.334	.051	15.555	0.000	1.000000
G25	27	4.70000	.334	.046	14.060	0.000	1.000000

Figure VI.20. Analysis of variance fit of 21 day lengths of surviving daphnids on beaker classification. LeBlanc test A.

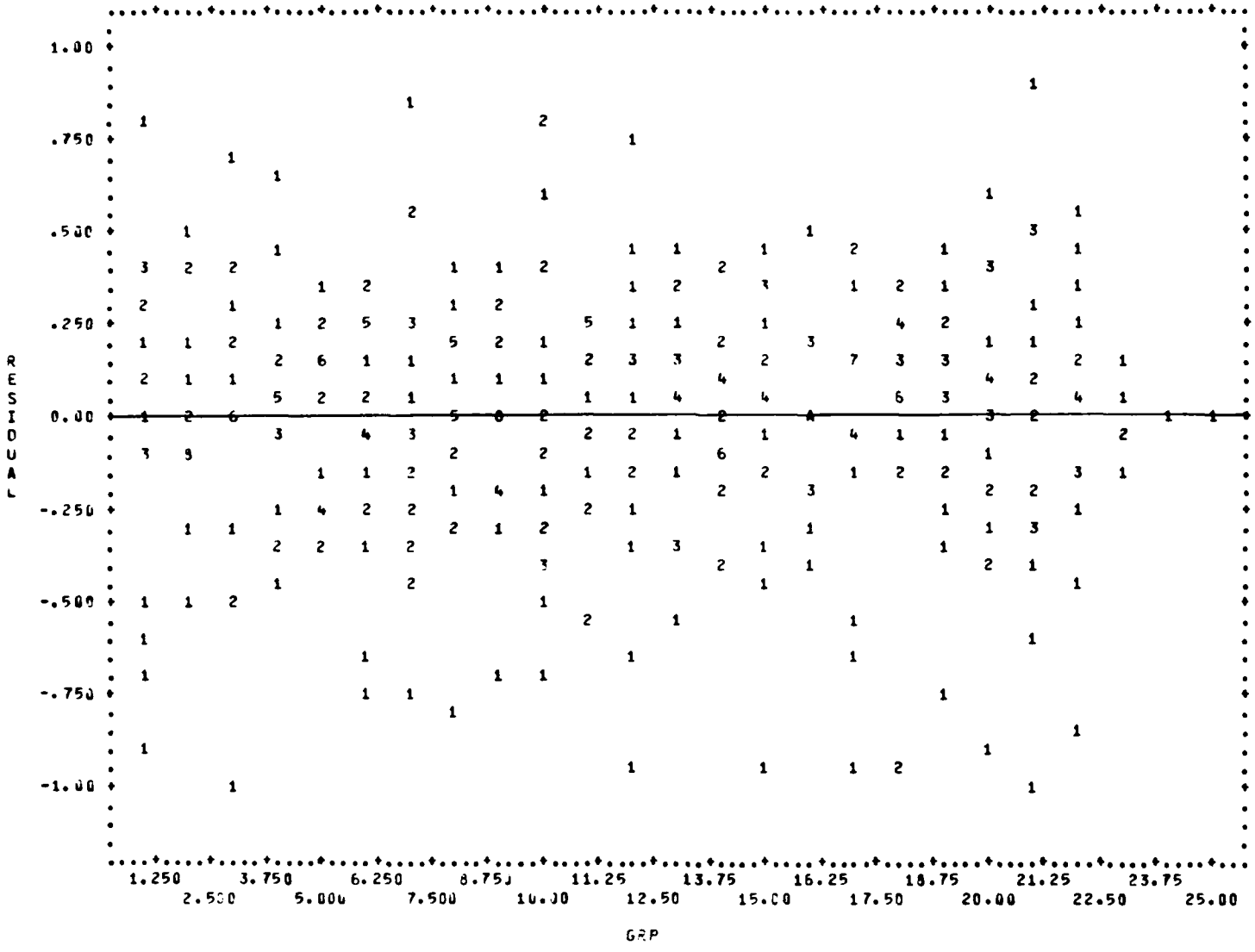


Figure VI.21. Residuals from analysis of variance fit versus beaker classification. LeBlanc test A.

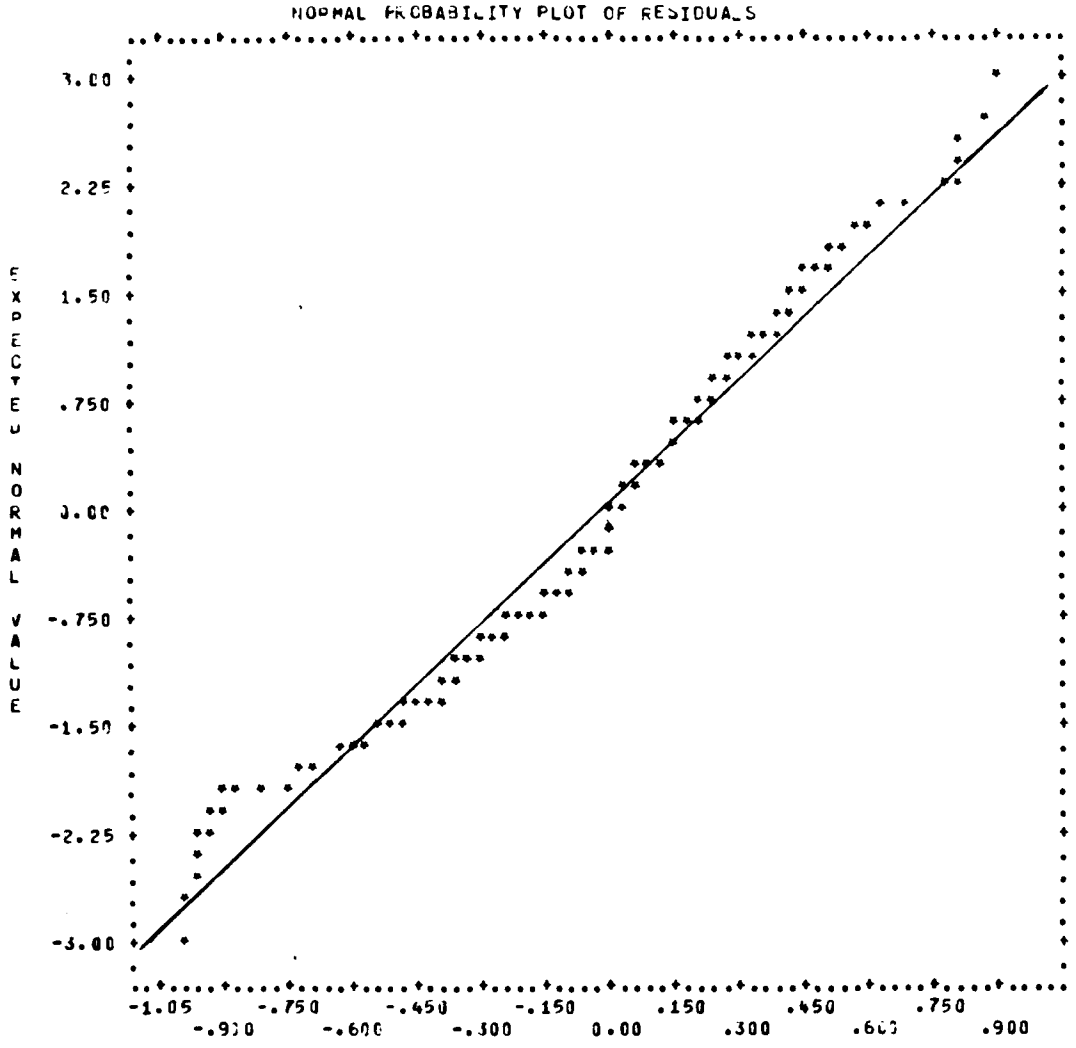


Figure VI.22. Normal probability plot of residuals from analysis of variance fit. LeBlanc test A.

REGRESSION TITLE. . . . . .BIONOMICS TEST B LENGTHS--PROB. PLOTS FOR TANK MEANS
DEPENDENT VARIABLE. . . . . 2 LENGTH
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R .9973 STD. ERROR OF EST. .3724
MULTIPLE R-SQUARE .9945

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	11675.673	28	388.417	2800.551	0.00000
RESIDUAL	59.777	431	.139		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
G1	3	5.02632	.085	.210	58.830	0.000	1.000000
G2	4	4.91111	.088	.199	55.949	0.000	1.000000
G3	5	4.91560	.083	.210	59.022	0.000	1.000000
G4	6	4.96000	.096	.184	51.582	0.000	1.000000
G5	7	4.80000	.090	.189	53.142	0.000	1.000000
G6	8	4.75000	.093	.182	51.018	0.000	1.000000
G7	9	4.65000	.093	.178	49.944	0.000	1.000000
G8	10	5.06667	.088	.236	57.721	0.000	1.000000
G9	11	4.79000	.083	.205	57.521	0.000	1.000000
G10	12	4.99444	.088	.203	56.898	0.000	1.000000
G11	13	4.67500	.083	.200	56.140	0.000	1.000000
G12	14	5.02778	.089	.204	57.278	0.000	1.000000
G13	15	4.96471	.090	.196	54.966	0.000	1.000000
G14	16	5.02000	.083	.215	60.283	0.000	1.000000
G15	17	4.93000	.083	.211	59.202	0.000	1.000000
G16	18	4.82222	.088	.196	54.936	0.000	1.000000
G17	19	4.83889	.088	.196	55.126	0.000	1.000000
G18	20	4.80556	.090	.189	53.207	0.000	1.000000
G19	21	4.77333	.096	.177	49.641	0.000	1.000000
G20	22	4.91765	.090	.194	54.445	0.000	1.000000
G21	23	4.93333	.088	.200	56.202	0.000	1.000000
G22	24	4.63000	.083	.198	55.599	0.000	1.000000
G23	25	4.80625	.093	.184	51.623	0.000	1.000000
G24	26	4.98333	.088	.202	56.771	0.000	1.000000
G25	27	4.56667	.215	.076	21.239	0.000	1.000000
G26	28	4.71429	.100	.169	47.364	0.000	1.000000
G27	29	4.80833	.100	.159	44.726	0.000	1.000000
G28	30	4.40000	.372	.042	11.815	0.000	1.000000

Figure VI.23. Analysis of variance fit of 21 day lengths of surviving daphnids on beaker classification. LeBlanc test B.

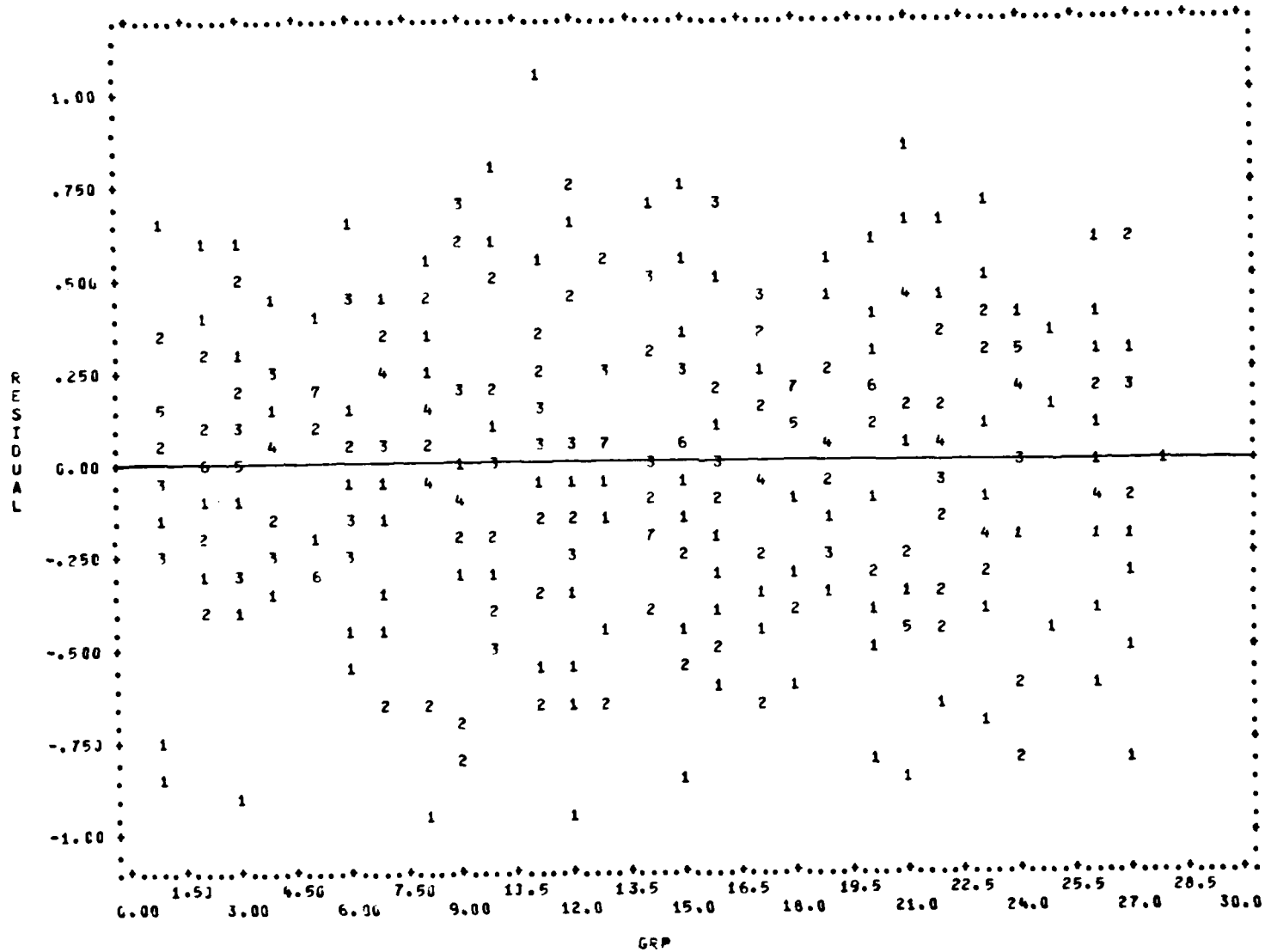


Figure VI.24. Residuals from analysis of variance fit versus beaker classification. LeBlanc test B.

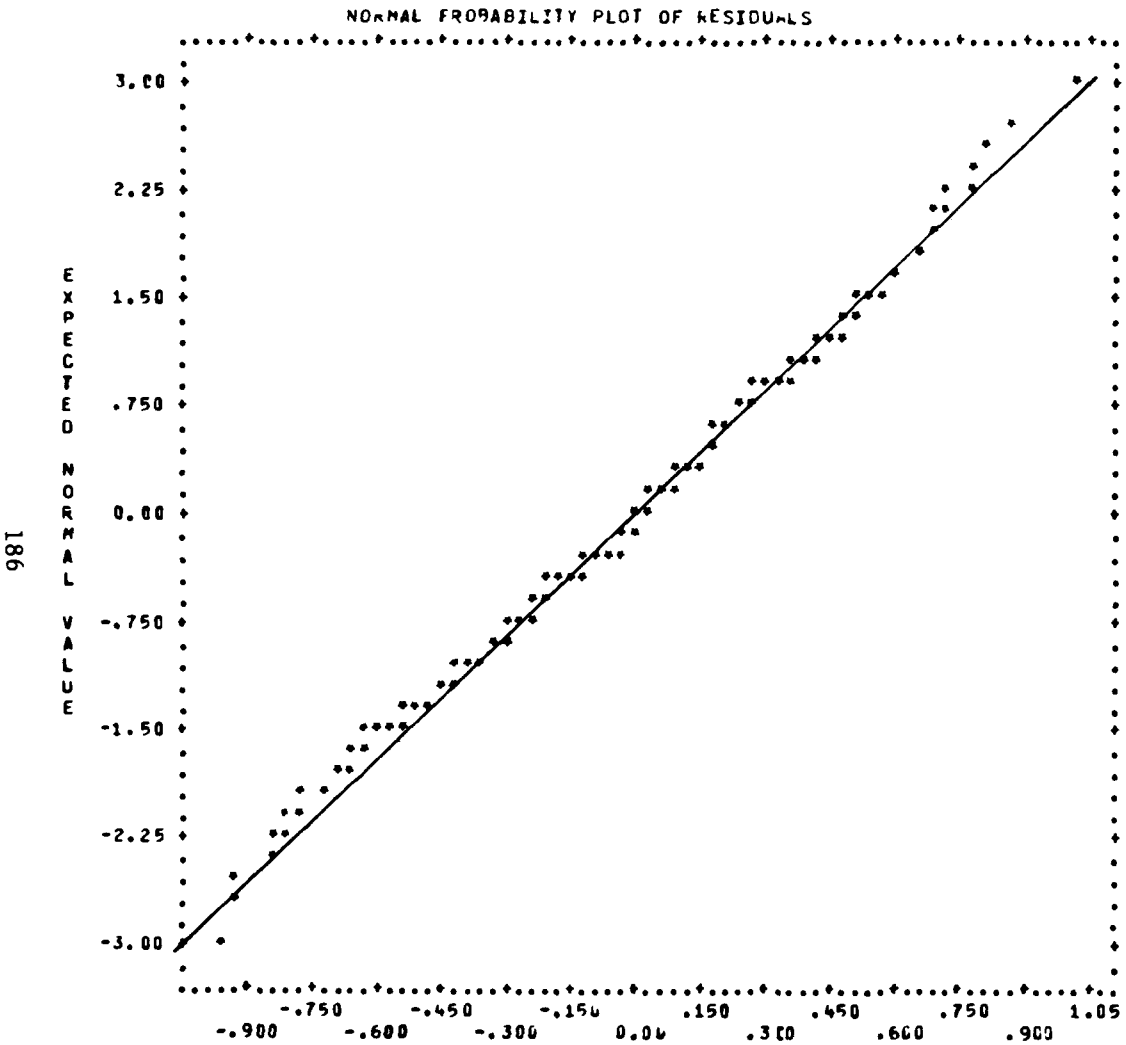


Figure VI.25. Normal probability plot of residuals from analysis of variance fit. LeBlanc test B.

REGRESSION TITLE. . . . . .CHAPMAN TEST--PROB. PLOTS FOR LENGTH
DEPENDENT VARIABLE. . . . . 12 LENGTH
TOLERANCE 0100

ALL DATA CONSIDERED AS A SINGLE GROUP

*** SEE WRITEUP FOR ZERO INTERCEPT RESULTS ***

MULTIPLE R	.9970	STD. ERROR OF EST.	.3280
MULTIPLE R-SQUARE	.9941		

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	994,622	8	124,328	1155,397	0,00000
RESIDUAL	5,918	55	,108		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
GRP1	14	4.28000	.104	.428	41.260	0.000	1.000000
GRP2	15	4.27500	.116	.382	36.861	0.000	1.000000
GRP3	16	4.12500	.116	.369	35.567	0.000	1.000000
GRP4	17	4.00000	.134	.310	29.869	0.000	1.000000
GRP5	18	3.81667	.134	.296	28.500	0.000	1.000000
GRP6	19	3.90000	.109	.370	35.667	0.000	1.000000
GRP7	20	3.59000	.104	.359	34.608	0.000	1.000000
GRP8	21	3.65000	.134	.283	27.255	0.000	1.000000

Figure VI.26. Analysis of variance fit of 21 day lengths of surviving daphnids on group. Chapman beryllium test.

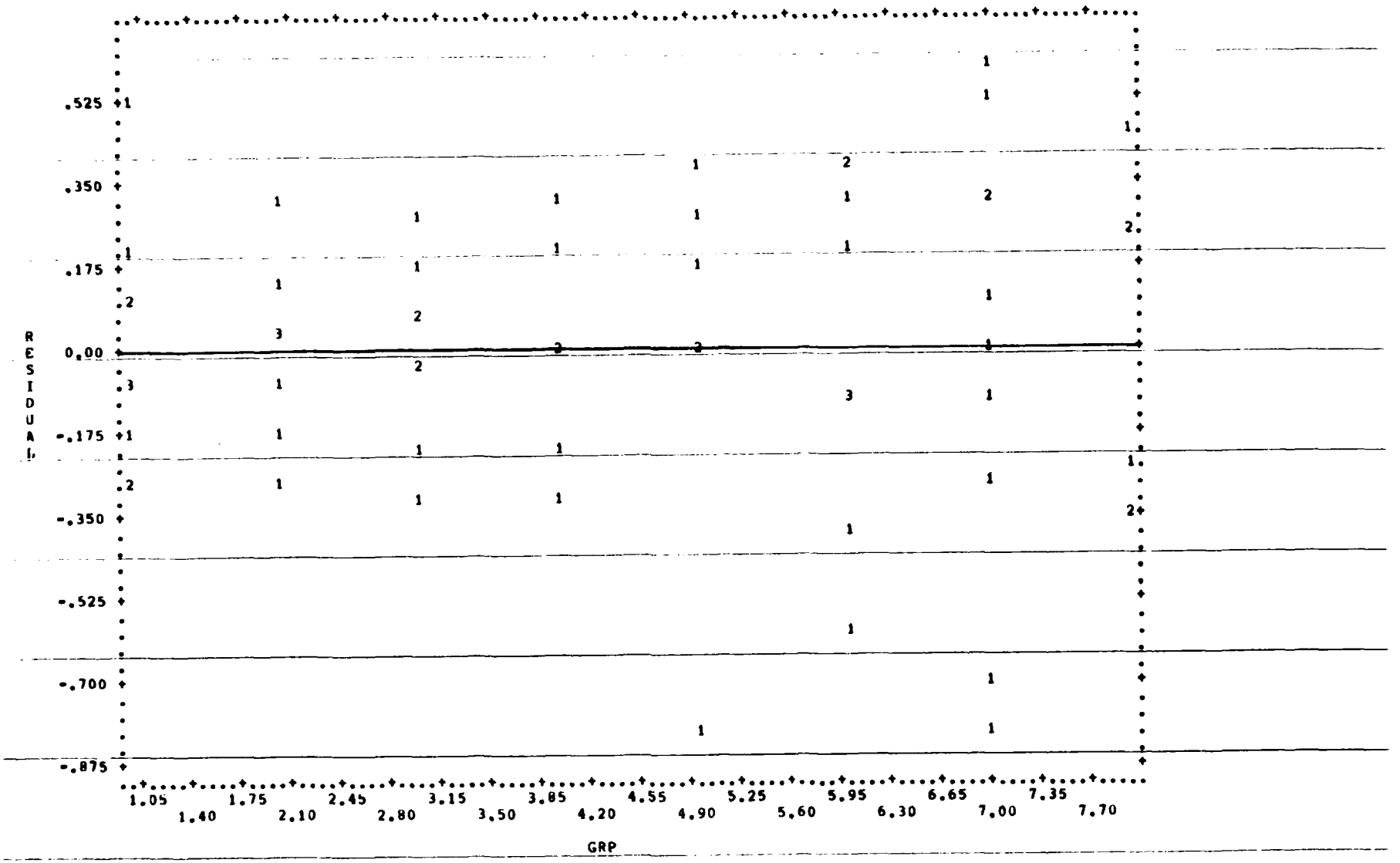


Figure VI.27. Residuals from analysis of variance fit versus group. Chapman beryllium test.

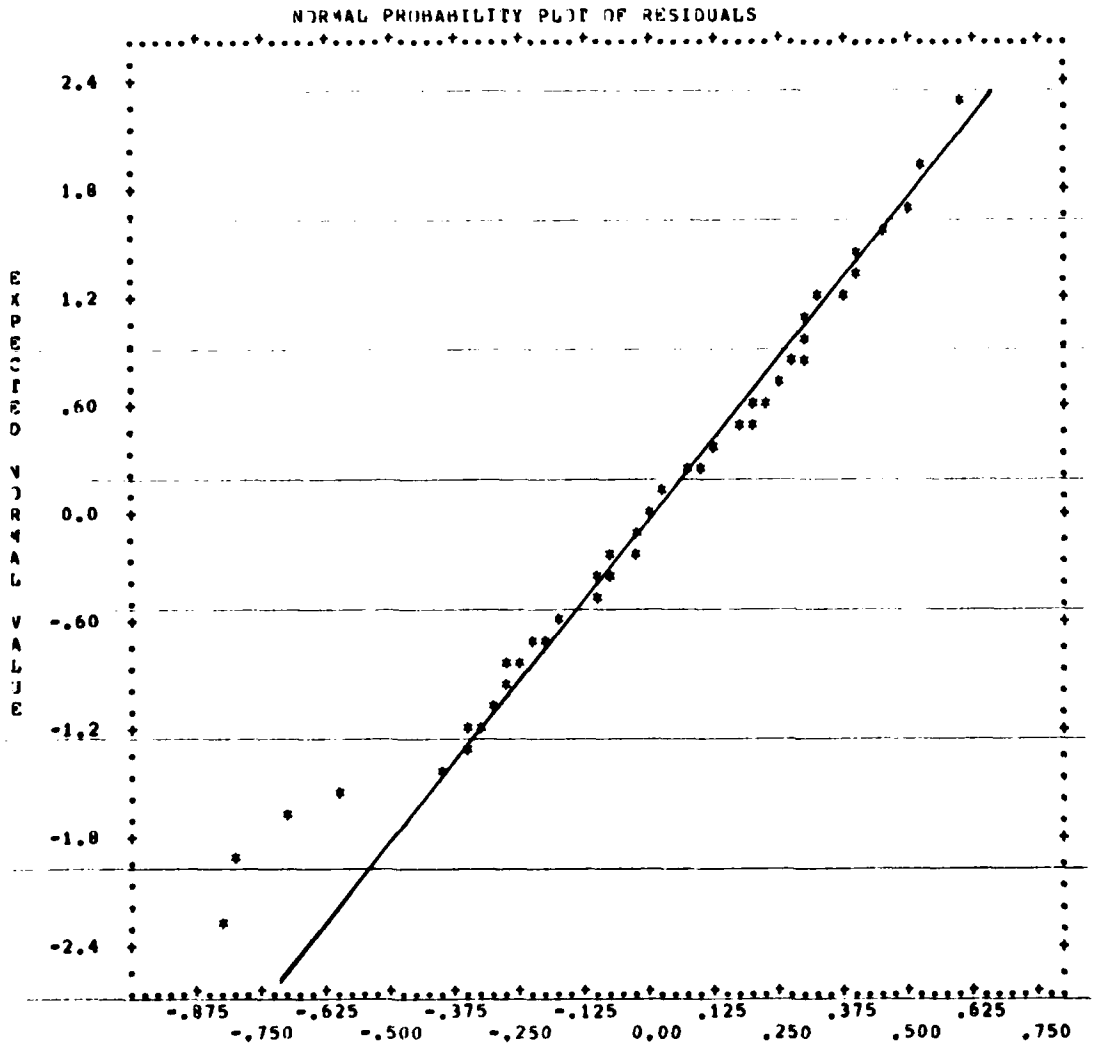


Figure VI.28. Normal probability plot of residuals from analysis of variance fit. Chapman beryllium test.

VII. COMPARISON OF WATER CONTROL AND SOLVENT CONTROL GROUPS

A. INTRODUCTION AND BACKGROUND

The toxicant under study may not be water soluble or may be soluble only at concentrations lower than those used in the test. In such a situation the toxicant must be dissolved in a solvent to prevent it from precipitating out of the water during the course of exposure. The substance under study is thus a combination of the toxicant of interest and the solvent used. The usual method of procedure is to prepare a relatively highly concentrated toxicant-solvent solution and utilize successive dilutions with water to arrive at the appropriate test concentrations. Thus the solvent is diluted along with the toxicant among the various test groups.

The utilization and successive dilution of solvents complicates the interpretation of toxic responses observed in a test. Since the solvent and toxicant are paired and diluted together, there is no way to sort out whether the observed effects are due to the toxicant, to the solvents, or to the pair. There is really no way to infer, solely on the basis of tests utilizing solvents, how the toxicant would act in the field in the absence of a solvent. (Note that concentrations above solubility levels can occur in the field due to effluents or spills.)

A partial solution to this dilema can be obtained by studying the effects of the solvent alone. If the solvent by itself produces no toxic responses at the concentrations utilized in the test, the assumption is made that any toxic effects observed in the test can be attributed to the toxicant. This assumption may or may not be valid in any given situation. The effects of the toxicant and the solvent may superimpose upon one another or even interact. The determination of the presence or extent of such joint effects would require special studies and special statistical analyses.

Since data from tests to study joint solvent-toxicant effects are not available and since exploration of this question is beyond the scope of this project, we content ourselves with a much more limited treatment of the problem. Namely, preliminary statistical tests are carried out to compare the average survival, length, and reproduction responses between the solvent and water control groups. If no statistically significant differences are found then we act as if there are no differences and we pool data across control groups. The combined responses are used for comparison with the treatment group responses. If statistically significant differences are found then treatment group responses are compared either just to the solvent control group responses or else separately to both the water control group and to the solvent control group responses. Although this is a commonly used approach to dealing with this question, it leaves much to be desired and raises a number of important conceptual issues.

First of all, the presence or absence of statistically significant differences in responses between water and solvent control groups is not the same thing as the presence or absence of biologically significant differences in responses between these groups. Statistical significance is a function of sample size as well as magnitude of effect. Thus the absence of a statistically significant difference between those groups does not imply that no difference exists. A biologically important difference might be revealed with use of greater numbers of daphnids or beakers.

Secondly, if the solvent is shown to have an effect, as evidenced by (statistically significant) differences between the solvent control group and water control responses, then even greater conceptual issues arise. Namely the toxicant effects and the solvent effects are confounded and there is no way to sort them out. The interpretational problem is further complicated if responses in certain treatment groups are significantly (biologically or statistically) different from one control group response but not from the other. Which comparison should be utilized? We side step the problem in this report by comparing both to the solvent and the water control groups if differences exist. We then note any differences in conclusions. However this does not answer the fundamental question of which comparison is more appropriate for regulatory or reporting purposes. That question cannot be answered solely on statistical grounds based on the test data available. Biological judgement must be used and/or additional tests must be carried out to study the effects of the solvent by itself. For our purposes we simply note the differences between the two control group responses, carry out separate comparisons with each control group, and remark that the results of the test are somewhat ambiguous. Any further interpretation would require additional testing and/or biological judgement.

Among the data sets considered in this report, LeBlanc's Tests A and B and Chapman's Beryllium Test utilize both solvent and water control groups. As will be seen below, significant differences between the solvent and water control groups in LeBlanc's Test A arise with respect to the survival, reproduction, and length responses. Subsections B, C and D below pertain to comparisons between the control groups in LeBlanc's Tests A and B and in Chapman's test respectively.

B. LE BLANC TEST A

Survival

The mortality rates in the water control group (group 1) and in the solvent control group (group 2) are compared by carrying out a 2 by 2 contingency table test for homogeneity utilizing the BMDP program, BMDP1F. The Pearson chi square test, with and without the continuity correction, and Fisher's exact test are reported. Two tailed tests are used. The output from these analyses is shown in Figure VII.1.

For the purpose of simplicity, analyses were first carried out based on the original (i.e. unadjusted for beaker to beaker heterogeneity) responses. If these comparisons show no differences, then comparisons based on the

adjusted responses surely would not. If significant differences are found, then we must redo the analyses after adjusting for beaker to beaker heterogeneity (based on the adjustments in Section V.)

We see from Figure VII.1 that there was 15 percent mortality in the water control group (12 of 80) and just 3.75 percent mortality in the solvent control group (3 of 80). Under the hypothesis of homogeneous mortality rates in the two groups, the minimum expected cell frequency is 7.5, thus suggesting that asymptotic theory is reasonable. The uncorrected chi square statistic is significant at the $\alpha = 0.0146$ level while the corrected chi square statistic is significant at the $\alpha = 0.0300$ level, in good agreement with Fisher's exact test. There is thus statistical evidence of differences in mortality rates between the two control groups, the solvent control group having a more favorable rate than the water control group.

The above tests did not account for the presence of beaker to beaker heterogeneity, which was shown in Section V to exist. From Table V.3 we see that the effective sample sizes and numbers of responses in groups 1 and 2, after adjustment for the effects of beaker to beaker heterogeneity, are:

Group	Dead	Live	Total
1	8.8	50.4	59.2
2	2.2	57.0	59.2
	11.0	107.4	118.4

The uncorrected chi square statistic is $\chi^2_u = 4.37$, which is significant at the $\alpha = 0.041$ level. The Yates corrected chi square statistic is $\chi^2_c = 3.14$, which is significant at the $\alpha = 0.081$ level. Since we associate 30 d.f. for error with each group (see Table V.3), the "chi square" statistics are compared to the percentiles of an F-distribution with degrees of freedom 1 and 60.

We thus conclude that there is statistical evidence of differences in mortality rates between the water control and solvent control groups. The solvent control group has lower mortality than the water control group.

Reproduction

The 21 day cumulative offspring per surviving adult in the water control group (group 1) and in the solvent control group (group 2) are compared either by carrying out a two sample, two tailed t-test or equivalently by a one way analysis of variance with two groups. Output from the latter analysis, utilizing the SPSS ANOVA procedure, is shown in Figure VII.2. Recall that these responses are obtained for each beaker by accumulating daily the total number of offspring produced on that day divided by the number of daphnids alive on that day.

The analysis in Figure VII.2 is carried out based on an error yardstick calculated only from the results in the two control groups (and thus having 6 d.f.). On the basis of this analysis we see that there is very strong statistical evidence of differences in production rates in these two groups. The production rate in the solvent control group is about 50 percent greater than that in the water control group. Thus unless there is some systematic difference between the water control and solvent control daphnids either with respect to biological hardiness or experimental handling, it appears as if the solvent is associated with increased production.

Figure VI.3 shows the results of an analysis of variance fit to all the groups. On the basis of this fit an error mean square of 460.345 with 21 d.f. is estimated. This is somewhat greater than the mean square of 278.1667 with 6 d.f. estimated from the firt in Figure VII.2 (although not significantly greater at $\alpha = 0.10$). An additional test was carried out using the larger error mean square. The difference between the production rates in the solvent and the water control groups is still very strongly significant ($\alpha = 0.000$).

Length

The 21 day lengths of the surviving adults in the water control group and in the solvent control group are compared either by carrying out a two sample, two tailed t-test or equivalently by a one way analysis of variance with two groups. Output from the latter analysis, utilizing the SPSS ANOVA procedure, is shown in Figure VII.3.

A basic difference between the length and reproduction responses, at least for multiply housed daphnids, is that lengths are measured on a per daphnid basis while reproduction is determined on a per beaker basis. Thus comparisons of lengths could be carried out on a per beaker basis, with one degree of freedom per beaker, or on a per daphnid basis, with one degree of freedom per daphnid (there were a total of 145 surviving daphnids in the two control groups). It was shown in Subsection IV.B that there is strong statistical evidence of beaker to beaker variation in lengths. Thus a per daphnid analysis would not be appropriate without somehow adjusting for the beaker to beaker heterogeneity. A method for carrying out such adjustments is discussed in Subsection V.D.

The analysis in Figure VII.3 is carried out on a per beaker basis, based on an error yardstick calculated only from the results in the two control groups (and thus having 6 d.f.). On the basis of this analysis we see that there is very strong statistical evidence of differences in average lengths in these two groups. The average length in the solvent control group is greater than that in the water control group. Thus unless there is some systematic difference in hardiness or in handling between the water control and the solvent control daphnids, it appears as if solvent is associated with increased lengths.

A suggested variability estimate, for comparisons of average lengths among groups, is discussed in Subsection V.D. The estimate is based on the variability among beaker averages within groups (actually the mean

square between beakers within groups, as discussed in Subsection IV.B) with degrees of freedom based on pooling information from the variability among beakers within groups and the variability among daphnids within beakers. The suggested variability estimate is MSI = 0.2722 with 98 d.f. This variability estimate is normalized to be on a per daphnid basis. In order to apply to average lengths within beakers we must divide it by the sample size within beakers. The average number of surviving adults per beaker in the two groups is 145/8 = 18.125. Thus the error estimate, normalized to a per average basis, is 0.2722/18.125 = 0.0150. (This compares with 0.0138 in Figure VII.3.) An additional test was carried out using the alternative error mean square with 98 d.f. Using this error yardstick yields the same conclusions as in Figure VII.3, namely that there is very strong statistical evidence (α = 0.0005) of differences in average lengths between the two control groups.

C. LE BLANC TEST B

The comparisons between the solvent and water control groups are carried out in the same manner as those discussed in the previous subsection for LeBlanc Test A. The discussion here will thus be less detailed.

Survival

The mortality responses in the water control and in the solvent control groups are compared by carrying out a 2 by 2 contingency table test for heterogeneity. Two tailed tests are used. The output is shown in Figure VII.4.

The analysis in Figure VII.4 is based on the original responses, unadjusted for beaker to beaker heterogeneity. The conclusion, based on the chi square test with or without correction or based on Fisher's exact test is that there is no statistical evidence of differences in mortality rates between the solvent and water control groups (α = 0.349). Since adjusting for beaker to beaker heterogeneity would only increase the observed significance level, it will not change our conclusions. This adjustment is thus omitted.

Reproduction

The 21 day cumulative offspring per surviving adult in the water control and solvent control groups are compared in the same manner as was done for the Test A responses, utilizing either a two tailed t-test or a one way analysis of variance test. The output appears in Figure VII.5. The analysis in Figure VII.5 is carried out based on an error yardstick calculated only from the results in the two control groups (and thus having 6 d.f.). On the basis of this analysis we conclude that there is no statistical evidence of differences in production rates in the two groups.

Figure VI.6 shows the results of an analysis of variance fit to all the groups. On the basis of this fit an error mean square of 461.786 with 21 d.f. is estimated. If we recalculate the F-ratio using this error yardstick we obtain

$$F = 40.50/461.786 = 0.088 \quad .$$

When compared to an F-distribution with degrees of freedom 1 and 21, the observed significance level is $\alpha = 0.770$. Thus, the conclusions are unchanged.

Length

The 21 day lengths of the surviving adults in the water control and solvent control groups are compared by either carrying out a two sample, two tailed t-test or equivalently by a one way analysis of variance with two groups. Output from the latter analysis, utilizing the SPSS ANOVA procedure, is shown in Figure VII.6.

It was shown in Subsection IV.B that there is strong statistical evidence of beaker to beaker variation in lengths. Thus the analysis in Figure VII.6 is carried out on a per beaker basis rather than on a per daphnid basis. Using the error yardstick calculated only from the results in the two control groups (and thus having 6 d.f.) we see that the average lengths in the two groups are not statistically significantly different.

A suggested variability estimate for comparisons of average lengths among groups, is discussed in Subsection V.D. The estimate is based on the mean square between beakers within groups with degrees of freedom based on pooling information from the variability among beakers within groups and the variability among daphnids within beakers. The suggested variability estimate is MSI = 0.2707 with 141 d.f. In order for the variability estimate to apply to average lengths within beakers we must divide it by the sample size within beakers. The average number of surviving adults per beaker in the two groups is 139/8 = 17.375. Thus the error estimate, normalized to a per average basis, is 0.2707/17.375 = 0.0156. (This compares with 0.0173 in Figure VII.6). Using this error yardstick we obtain the F-ratio

$$F = 0.0372/0.0156 = 2.391 \quad .$$

We compare this to the percentiles of an F-distribution with d.f. 1 and 141. The resulting significance level is $\alpha = 0.124$. This is at most marginal. There is thus a suggestion that the average lengths in the water control groups are greater than those in the solvent control group, but nothing conclusive. We will combine the responses from both control groups in subsequent analyses.

It should be noted that the variability among beaker averages is significantly greater ($\alpha = 0.04$) in the solvent control group than in the water control group. Figure II.29 shows that the greater variability in the solvent control group is due to a single beaker average which is somewhat removed from the others in that group. Without this relatively high value, the average length in the solvent control group would be substantially lower than that in the water control group and there would probably be a statistically and biologically significant differences between them.

D. CHAPMAN - BERYLLIUM

Ten daphnids were placed on test in each group, one per beaker. Thus the sample sizes in this data set are considerably smaller than those in the LeBlanc data sets. Furthermore, since the daphnids were individually housed, there is no question of beaker to beaker variation producing correlated responses.

Survival

The mortality responses in the water control and in the solvent control groups are compared by carrying out a 2 by 2 contingency table test for heterogeneity. Two tailed tests are used. The output is shown in Figure VII.7.

Because of the small numbers of daphnids per groups, the expected frequencies are relatively small. Under the hypothesis of homogeneity, the expected number of dead daphnids per group is 1.0. This raises questions about the validity of the asymptotic theory on which the usual Pearson chi square test is based. We see, in fact, that the observed significance levels of Fisher's exact test (two tailed) and the chi square test with correction differ considerably from the observed significance level of the uncorrected chi square test. The conclusion, based on the chi square test with correction or based on Fisher's exact test, is that there is no statistical evidence of differences in mortality between the solvent and water control group ($\alpha = 0.47$).

Reproduction

Comparisons between water and solvent control groups are based only on those daphnids that survived to the end of the test. There were ten survivors in the water control group and eight survivors in the solvent control group. The daphnids that died early obviously present a distorted view of total number of young produced and so their responses are not included in this comparison. Total numbers of young per individual adult daphnid can be determined for these data since the daphnids were housed just one to a beaker.

The output from the comparison appears in Figure VII.8. The analysis in Figure VII.8 is carried out by means of a one way analysis of variance with two groups, based on an error yardstick calculated only from the results in the two control groups (and thus having 16 d.f.). On the basis of this analysis we conclude that there is no statistical evidence of differences in production rates in the two groups. However the average production rate in the solvent control group is about 18 percent higher than that in the water control group.

Figure VI.12 shows the results of an analysis of variance fit to all the groups. On the basis of this fit an error mean square of 2365.759 with 55 d.f. is estimated. This mean square is somewhat lower than that in Figure VII.8, based solely on the control groups. The plot in Figure II.17 suggests

that the control group responses are higher and more variable than the treatment group responses. (However a two tailed F-test to compare the error mean square in the treatment groups with that in the control groups is nonsignificant ($F = 1.71$, $\alpha = 0.17$)). If we recalculate the F-ratio in Figure VII.8 using the error estimate in Figure VI.12 we obtain

$$F = 3240.00/2365.759 = 1.369 \quad .$$

When compared to an F-distribution with degrees of freedom 1 and 55, the observed significance level is $\alpha = 0.247$. Thus, the conclusions are unchanged.

Length

The 21 day lengths of surviving adults in the water control and solvent control groups are compared by carrying out a two sample, two tailed t-test or equivalently by a one way analysis of variance with two groups. Output from the one way analysis of variance appears in Figure VII.9. The analysis in Figure VII.9 is based on an error yardstick calculated only from the results in the two control groups (and thus having 16 d.f.). On the basis of this analysis we conclude that there is no statistical evidence of differences in average lengths among survivors in the two groups. The average lengths are virtually identical (4.28 mm in the water control group and 4.275 mm in the solvent control group).

Figure VI.26 shows the results of an analysis of variance fit to all the groups. On the basis of this fit an error mean square of 0.108 with 55 d.f. is estimated. This mean square is somewhat greater than that in Figure VII.9, based solely on the control groups. (A two tailed F-test to compare the error mean square in the treatment groups with that in the control groups is significant at $\alpha = 0.04$). Thus recalculating the F-ratio based on the mean square in Figure VI.26 would not change the conclusions arrived at based on the analysis in Figure VII.9.

E. SUMMARY

Three of the data sets under consideration contain both water and solvent control groups. These are LeBlanc-Tests A and B and Chapman-Beryllium Test. Comparisons of the average survival, length and reproduction responses were made between the two control groups in each of the tests. Statistically (and perhaps biologically) significant differences were seen for all three responses in LeBlanc's Test A, especially for reproduction and length. No statistically significant differences were found for any of the responses either in LeBlanc's Test B or in Chapman's Beryllium Test.

Based on the outcomes of these comparisons, in subsequent sections we will base comparisons of treatment and control group responses on pooled water and solvent control group responses in LeBlanc's Test B data and in Chapman's Beryllium data and we will make separate comparisons with the water control and with the solvent control group responses in LeBlanc's Test A. The results of both sets of comparisons in LeBlanc Test A will be reported. However this will not resolve the conceptual issues of interpretation that were discussed in Subsection A.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 4)

TABLE NO. 1 GRP (VAR 2) VS SURVIVAL(VAR 3)

CELL FREQUENCY COUNTS

			SURVIVAL(VAR 3)		
			ALIVE	DEAD	
		EQ./EQ.	1.00	2.00	TOTAL
GRP	WATER	1.00	68	12	80
(VAR 2)	SOLVENT	2.00	77	3	80
		TOTAL	145	15	160

MINIMUM UNWEIGHTED EXPECTED VALUE IS 7.500
***** ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
FISHER EXACT TEST(1-TAIL)			.0136	FISHER EXACT TEST(2-TAIL)			.0271
PEARSON CHISQUARE	5.959	1	.0146	YATES' CORRECTED CHISQUARE	4.708	1	.0300

PERCENTAGES OF THE TOTAL FREQUENCY

			SURVIVAL(VAR 3)		
			ALIVE	DEAD	
		EQ./EQ.	1.00	2.00	TOTAL
GRP	WATER	1.00	42.50	7.50	50.00
(VAR 2)	SOLVENT	2.00	48.13	1.88	50.00
		TOTAL	90.63	9.38	100.00

PERCENTAGES OF THE ROW TOTALS

			SURVIVAL(VAR 3)		
			ALIVE	DEAD	
		EQ./EQ.	1.00	2.00	TOTAL
GRP	WATER	1.00	85.00	15.00	100.00
(VAR 2)	SOLVENT	2.00	96.25	3.75	100.00
		TOTAL	90.63	9.38	100.00

Figure VII.1. Comparison of mortality rates in solvent and water control groups. LeBlanc test A. (Group 1 - water control, Group 2 - solvent control).

BIONOMICS TESTS A&H--WATER VS SOLVENT CONTROLS FOR OFFSPRING 07/02/81 14.16.42. PAGE 2

FILE NONAME (CREATION DATE = 07/02/81)
SUBFILE TESTA

- O N E W A Y -

VARIABLE PERF21
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	5304.5000	5304.5000	19.070	.0047
WITHIN GROUPS	6	1669.0000	278.1667		
TOTAL	7	6973.5000			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	4	102.5000	19.2959	9.6480	84.0000	129.0000	71.7963 TO 133.2037
SOLVENT	4	154.0000	13.5647	6.7823	134.0000	164.0000	132.4159 TO 175.5841
TOTAL	8	128.2500			84.0000	164.0000	
		UNGROUPED DATA	31.5629	11.1592			101.8628 TO 154.6372

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .6693, P = .577 (APPROX.)
BARTLETT-BOX F = .312, P = .578
MAXIMUM VARIANCE / MINIMUM VARIANCE = 2.024

Figure VII.2. Comparison of cumulative offspring per surviving adult between solvent and water control groups. LeBlanc test A.

BIONOMICS TESTS A&M--WATER VS SOLVENT CONTROLS FOR LENGTHS 06/30/81 16.41.29. PAGE 2
USING TANK MEANS AS DATA--INDIVIDUALS AVERAGED WITHIN TANKS

FILE NONAME (CREATION DATE = 06/30/81)
SUBFILE TESTA

- O N E W A Y -

VARIABLE LENGTH
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	.1920	.1920	13.908	.0097
WITHIN GROUPS	6	.0828	.0138		
TOTAL	7	.2749			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	4	4.9676	.1555	.0778	4.8000	5.1000	4.7202 TO 5.2151
SOLVENT	4	5.2775	.0585	.0293	5.2300	5.3500	5.1844 TO 5.3706
TOTAL	8	5.1226			4.8000	5.3500	
	UNGROUPED DATA		.1982	.0701			4.9569 TO 5.2882

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .8760, P = .143 (APPROX.)
BARTLETT-BOX F = 2.172, P = .143
MAXIMUM VARIANCE / MINIMUM VARIANCE = 7.062

Figure VII.3. Comparison of 21 day lengths of surviving adults between solvent and water control groups. LeBlanc test A.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 4)

TABLE NO. 1 GRP (VAR 2) VS SURVIVAL(VAR 3)

CELL FREQUENCY COUNTS

			SURVIVAL(VAR 3)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP	WATER	1.00	72	8	80
(VAR 2)	SOLVENT	2.00	67	13	80
		TOTAL	139	21	160

MINIMUM UNWEIGHTED EXPECTED VALUE IS 10.500

ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
FISHER EXACT TEST(1-TAIL)			.1747	FISHER EXACT TEST(2-TAIL)			.3493
PEARSON CHISQUARE	1.370	1	.2418	YATES' CORRECTED CHISQUARE	.877	1	.3490

PERCENTAGES OF THE TOTAL FREQUENCY

			SURVIVAL(VAR 3)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP	WATER	1.00	45.00	5.00	50.00
(VAR 2)	SOLVENT	2.00	41.88	8.13	50.00
		TOTAL	86.88	13.13	100.00

PERCENTAGES OF THE ROW TOTALS

			SURVIVAL(VAR 3)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP	WATER	1.00	90.00	10.00	100.00
(VAR 2)	SOLVENT	2.00	83.75	16.25	100.00
		TOTAL	86.88	13.13	100.00

Figure VII.4. Comparison of mortality rates in solvent and water control groups. LeBlanc test B. (Group 1 - water control, Group 2 - solvent control).

BIONOMICS TESTS ACH--WATER VS SOLVENT CONTROLS FOR OFFSPRING   07/02/81   14.16.42.   PAGE 3

FILE NONAME (CREATION DATE = 07/02/81)
SUBFILE CHST3

- O N E W A Y -

VARIABLE FERT21
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	40.5000	40.5000	.092	.7720
WITHIN GROUPS	6	2643.0000	440.5000		
TOTAL	7	2683.5000			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	4	111.5000	24.0069	12.0035	79.0000	132.0000	73.3002 TO 149.6998
SOLVENT	4	107.0000	17.4547	8.7274	83.0000	123.0000	79.2261 TO 134.7739
TOTAL	8	109.2500			79.0000	132.0000	
		UNGROUPED DATA	19.5795	6.9224			92.8812 TO 125.6188

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .6542, P = .614 (APPROX.)
BARTLETT-BOX F = .256, P = .614
MAXIMUM VARIANCE / MINIMUM VARIANCE = 1.892

Figure VII.5. Comparison of cumulative offspring per surviving adult between solvent and water control groups. LeBlanc test B.

BIOASSAYS TESTS ABO--DATA VS SOLVENT CONTROLS FOR LENGTHS 06/30/81 16.41.29. PAGE 3
USING TANK MEANS AS DATA--INDIVIDUALS AVERAGED WITHIN TANKS
FILE NEWAYS (CREATION DATE = 06/30/81)
SPECIFY TESTS

- O N E W A Y -

VARIABLE LENGTH
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	.0372	.0372	2.156	.1924
WITHIN GROUPS	6	.1036	.0173		
TOTAL	7	.1409			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	4	4.9531	.0515	.0258	4.9111	5.0263	4.8678 TO 5.0384
SOLVENT	4	4.8167	.1780	.0890	4.6500	5.0667	4.5335 TO 5.0998
TOTAL	8	4.8849			4.6500	5.0667	
	UNGROUPED DATA		.1419	.0502			4.7663 TO 5.0035

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .9168, P = .079 (APPROX.)
BARTLETT-BOX F = 3.121, P = .080
MAXIMUM VARIANCE / MINIMUM VARIANCE = 11.017

Figure VII.6. Comparison of 21 day lengths of surviving adults between solvent and water control groups. LeBlanc test B.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 3)

TABLE NO. 1 GRP (VAR 1) VS SURVIVAL(VAR 2)

CELL FREQUENCY COUNTS

		SURVIVAL(VAR 2)		
	EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	WATER 1.00	10	0	10
	SOLVENT 2.00	8	2	10
	TOTAL	18	2	20

MINIMUM EXPECTED VALUE IS 1.000

ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
FISHER EXACT TEST(1-TAIL)			.2368	FISHER EXACT TEST(2-TAIL)			.4737
PEARSON CHISQUARE	2.222	1	.1360	YATES' CORRECTED CHISQUARE	.556	1	.4561

PERCENTAGES OF THE TOTAL FREQUENCY

		SURVIVAL(VAR 2)		
	EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	WATER 1.00	50.00	0.00	50.00
	SOLVENT 2.00	40.00	10.00	50.00
	TOTAL	90.00	10.00	100.00

PERCENTAGES OF THE ROW TOTALS

		SURVIVAL(VAR 2)		
	EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	WATER 1.00	100.00	0.00	100.00
	SOLVENT 2.00	80.00	20.00	100.00
	TOTAL	90.00	10.00	100.00

Figure VII.7. Comparison of mortality rates in solvent and water control groups. Chapman - beryllium. (Group 1 - water control, Group 2 - solvent control).

CHAPMAN TEST--WATER VS SOLVENT CONTROLS FOR LENGTHS, OFFSPRING 06/30/81 16.17.44. PAGE 3

FILE NONAME (CREATION DATE = 06/30/81)

- O N E W A Y -

VARIABLE TOTALYIG - FOR ADULTS SURVIVING TO THE END OF THE TEST
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	3240.0000	3240.0000	.967	.3401
WITHIN GROUPS	16	53622.5000	3351.4062		
TOTAL	17	56862.5000			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	10	150.5000	67.0278	21.1960	65.0000	252.0000	102.5512 TO 198.4488
SOLVENT	8	177.5000	43.4051	15.3460	99.0000	239.0000	141.2125 TO 213.7875
TOTAL	18	162.5000			65.0000	252.0000	
	UNGROUPED DATA		57.8347	13.6318			133.7395 TO 191.2605

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .7046, P = .240 (APPROX.)
BARTLETT-BOX F = 1.311, P = .253
MAXIMUM VARIANCE / MINIMUM VARIANCE = 2.385

Figure VII.8. Comparison of total number of offspring between solvent and water control groups. Chapman - beryllium. Comparison is confined to those daphnids that survived to the end of the test.

CHAPMAN TEST--WATER VS SOLVENT CONTROLS FOR LENGTHS, OFFSPRING 06/30/81 16.17.44. PAGE 2

FILE NONAME (CREATION DATE = 06/30/81)

- O N E - W A Y -

VARIABLE LENGTH
BY GRP

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	.0001	.0001	.002	.9628
WITHIN GROUPS	16	.7910	.0494		
TOTAL	17	.7911			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN
WATER	10	4.2800	.2486	.0786	4.0300	4.8000	4.1022 TO 4.4578
SOLVENT	8	4.2750	.1832	.0648	4.0200	4.6000	4.1218 TO 4.4282
TOTAL	18	4.2778			4.0200	4.8000	
	UNGROUPED DATA		.2157	.0508			4.1705 TO 4.3851

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRANS C = MAX.VARIANCE/SUM(VARIANCES) = .6479, P = .407 (APPROX.)
BARTLETT-BOX F = .661, P = .416
MAXIMUM VARIANCE / MINIMUM VARIANCE = 1.840

Figure VII.9. Comparison of 21 day lengths of surviving adults between solvent and water control groups. Chapman - beryllium.

VIII. COMPARISON OF LENGTH AND REPRODUCTION RESPONSES

A. INTRODUCTION

A number of toxicologists have informally suggested that there is an association between the lengths of adult daphnids and the numbers of offspring that they produce. If there is in fact a strong positive correlation between these variables then perhaps it would suffice just to measure the adult daphnids rather than to count their offspring. The latter task is much more tedious and time consuming than the former and so the possibility of being able to do away with measuring reproduction is attractive from both the cost and work standpoint. However before suggesting this, we would need to determine whether any information would be lost by relying solely on length measurements.

Some individuals have even gone a step further (with tongue firmly implanted in cheek?) and suggested that there is a strong association between 7 day lengths and 21 day lengths, so that adult daphnids need only be measured on day 7. Thus to carry matters to the extreme, life cycle effects could be predicted by 21 day lengths and reproduction, which could be predicted by 21 day lengths, which in turn could be predicted by 7 day lengths! That would introduce wonderful economics of operation!

We limit the discussion in this section to briefly studying the association between 21 day lengths and 21 day cumulative reproduction per survivor. Subsection B contains scatterplots of the relations between 21 day lengths and reproduction. Subsection C contains regression fits relating these variables, plots of residuals from the fits, and inferences about the predictability of cumulative reproduction based on lengths. Subsection D contains a summary and conclusions.

B. PRELIMINARY SCATTERPLOTS

This subsection contains scatterplots that display the associations between 21 day lengths and 21 day cumulative reproduction for LeBlanc's Tests A and B and for Chapman's Beryllium Test. In both of LeBlanc's data sets the daphnids were multiply housed, 20 to a beaker at the outset of the test. Thus reproduction was determined on a per beaker basis and cannot be associated with individual daphnids. By contrast lengths were determined on a per daphnid basis and so can be associated with individual daphnids. Thus the only comparisons that can be made in LeBlanc's data sets are on a beaker basis. Namely 21 day cumulative production per surviving adult is compared to 21 day average length among survivors, for each beaker. Of course the reproduction and length values are not necessarily based on exactly the same set of daphnids since those that died before the end of the test influence the reproduction values but not the average lengths. In Chapman's data the situation is better. Since the daphnids were housed individually, 21 day cumulative production and 21 day length determinations can be associated on a daphnid by daphnid basis.

Figure VIII.1 shows the relationship between 21 day cumulative reproduction per surviving adult and average 21 day lengths for LeBlanc's Test A. Each point corresponds to a beaker and plotting symbol is the group number. Note that the results from group 7 have been excluded from this plot because of high mortality. Two or more coincident points from different groups are plotted as an asterisk. There is a definite positive association between these responses. The trend appears approximately linear for higher length values and flattens out for lower length values. There is a fair bit of scatter about the general trend.

Figure VIII.2 shows the relationship between 21 day cumulative reproduction per surviving adult and average 21 day lengths for LeBlanc's Test B. The plotting symbols have the same meaning as in Figure VIII.1. There is no association between the two responses in this test.

Figure VIII.3 shows the relationship between 21 day cumulative reproduction and 21 day lengths for daphnids that survived to the end in Chapman's Beryllium Test. Each point corresponds to an individual daphnid. Plotting symbol is the group number. Two or more coincident points from different groups are plotted as an asterisk. There is a rather strong positive association between the responses, stronger than that in Figure VIII.1 for LeBlanc's data. The trend appears approximately linear for higher length values and flattens out for lower length values. There is much less scatter about this trend than about the trend in Figure VIII.1.

In order to enhance the straight line trend in the lower portion of the plot, a logarithmic transformation of cumulative production was made. Figure VIII.4 shows the relationship between $\log_{10}$ (21 day cumulative production) and 21 day lengths for daphnids that survived to the end of the test. There now appears to be a straight line trend throughout the entire range of length values, however the variability about the trend line is greater for shorter daphnids than for taller ones. The strength of the relationship is again very much greater than that in Figure VIII.1.

C. REGRESSION ANALYSES

Based on the appearances of the preliminary scatterplots in Subsection B it appears that there is at least some association between length and cumulative reproduction. To quantify the extent of this relationship, simple linear regression models were fitted to predict production from length in LeBlanc's Tests A and B and in Chapman's Beryllium Test and to predict $\log_{10}$ (production) from length in Chapman's Beryllium Test. The results of these regression fits are shown in Figures VIII.5 to VIII.8.

The results in Figure VIII.5 show that there is a significant association between length and reproduction but lengths explain just 29 percent of the variation in production. Thus better than two thirds of the variation in production is explained by factors other than length. This and the appearance of Figure VIII.1 strongly suggest that 21 day average length per beaker is not in and of itself a good predictor of 21 day cumulative reproduction in Test A, although the two responses are related.

The results in Figure VIII.6 and the scatterplot in Figure VIII.2 show no evidence of any association between 21 day average length per beaker and 21 day cumulative mortality in Test B.

The results in Figures VIII.7 and VIII.8 and the scatterplots in Figures VIII.3 and VIII.4 show moderately strong associations between 21 day length and 21 day cumulative production per daphnid in Chapman's Beryllium Test. The regression in Figure VIII.8 explains nearly 60 percent of the variability in $\log_{10}$ (production). This is a much stronger association than was the case in LeBlanc's tests. It is interesting to note that the parameters of the regression model fitted to the Test A data in Figure VIII.5 are quite similar to those fitted to the Beryllium data in Figure VIII.7.

In summary, the degree of association is strongest in Chapman's test and weakest in LeBlanc's Test B. The degree of association differs markedly from test to test.

D. RESIDUAL DISPLAYS

Residuals are the differences between the observed values of the responses and the values predicted by the regression model. The residuals reveal systematic structure in the data that was not accounted for by the regression fit, departures from model assumptions such as lack of independence or nonconstant variance, outliers in the data, associations with variables not included in the fit, etc. Residuals can be studied by preparing various types of graphical and numerical displays that look for particular types of structure. If the fitted regression model accounts for all the systematic behavior in the responses, the residuals should resemble random noise. Any systematic behavior observed in the residuals suggests that the fitted regression model is not fully adequate to describe all the structure in the data.

Plots of residuals versus predicted values, squares of residuals versus predicted values, normal probability plots of residuals, and scatterplots of residuals versus group number were prepared. Most of the plots did not reveal any anomalies or departures from assumptions. However the plots of residuals by group revealed some systematic structure in the data over and above that explained by the regression model. These plots are shown in Figures VIII.9 to VIII.11. If the lengths accounted for all the systematic behavior in production then the residuals in these plots should be distributed randomly about 0 with no trends across groups. This is seen not to be the case.

Figure VIII.9 displays the residuals from the regression fit in LeBlanc's Test A plotted by group. Recall that the group 7 results were deleted because of the very high mortality rate. A systematic pattern can be seen. Namely there is a generally upward trend in the residuals as group number increases from 1 to 5 and then a sharp drop in the residuals in group 6. This suggests that average length decreases more rapidly than production in groups 1-5 and production decreases much more rapidly than length in

going from group 5 to group 6. Since group number is really a surrogate for concentration level, this behavior suggests that toxicant concentration impacts cumulative production in a manner that cannot be fully explained by its impact on length. That is, concentration is associated with production over and above the association of length with production.

Figure VIII.10 displays the residuals from the regression fit in LeBlanc's Test B plotted by group. A systematic pattern is again evident. There is a curvilinear trend in the average residual within groups, first rising and then falling. This suggests that for lower concentrations length decreases more rapidly than productivity while for higher concentrations productivity decreases mor rapidly than length. This again shows that productivity is associated with concentration over and above its association with length.

Figure VIII.11 displays the residuals from the regression fit in Chapman's test on beryllium plotted by group. A systematic trend is again evident. There is a curvilinear trend in the average residual within groups, first falling and then rising. This trend may be due to lack of fit of the simple linear regression model or to effects of concentration on production that are not reflected in effects on length. To assess whether there is any systematic lack of fit in the simple linear regression relating length and production, the residuals from this fit were plotted versus predicted values (which are essentially proportional to length). The plot is displayed in Figure VIII.12.

To assess whether there is any systematic trend in the residuals, the range of predicted values was subdivided so that there would generally be 10-20 points within each interval. The median of the residuals was calculated in each interval and is indicated by the symbol ⊠. The trend in these medians is characteristic of trends that reflect departures from polynomial models. It is interesting to note that the trend in Figure VIII.11 mimics that in Figure VIII.12, although the extent of the dip in the group means in Figure VIII.11 is deeper than the dip in the medians in Figure VIII.12. Thus at least part of the trend observed in Figure VIII.11 might be eliminated by adding quadratic or cubic terms to the regression in Figure VIII.8 relating $\log_{10}$ (reproduction) and length. This should be done to determine if it improves the fit and the residuals should be recalculated. Since this path was not pursued, we cannot comment about the extent to which the trend in Figure VIII.11 reflects quadratic or cubic effects in the regression relationship between length and $\log_{10}$ (reproduction) or the extent to which it reflects effects of concentration on production not associated with effects on length. Sorting this out will await future work. However the nearly random appearance of Figure VIII.12 suggests that adding nonlinear terms to the regression model will not markedly improve the strength of the fit. Thus most of the systematic behavior in Figure VIII.11 is probably due to effects of factors other than length.

E. INFERENCES BASED ON THE REGRESSION MODEL

The adequacy or inadequacy of a regression fit depends on whether inferences of interest can be made with sufficient precision to be of practical use. The natural type of inference to be made from a regression model relating length and production is the prediction of expected production given length. This might be in the form of a confidence interval on mean production conditional on length or a prediction interval on the sample average of say 10 daphnids, conditional on length. We consider the calculation of 95 percent confidence intervals for various values of length. We illustrate the procedure with Chapman's beryllium data. Let LPROD denote $\log_{10}$ (PROD). From Figure VIII.8 we determine that the fitted regression model is

$$\hat{\text{LPROD}} = \hat{\beta}_0 + \hat{\beta}_1 \text{ LENGTH} \equiv 0.15515 + 0.46306 \text{ LENGTH} \quad .$$

Now the average length is 3.96508 mm. Thus the above equation can be rewritten as

$$\hat{\text{LPROD}} = \overline{\text{LPROD}} + \hat{\beta}_1 \text{ (LENGTH} - \overline{\text{LENGTH}}) = 1.9912 + 0.46306 \text{ (LENGTH} - \overline{\text{LENGTH}}) \quad .$$

Thus the standard error of $\hat{\text{LPROD}}$ is

$$[\hat{\sigma}^2/n + \hat{\sigma}^2 \, (\hat{\beta}_1) \text{ (LENGTH} - \overline{\text{LENGTH}})^2]^{1/2} \equiv \hat{\sigma} \, (\hat{\text{LPROD}}) =$$

$$[0.0241/63 + 0.0024 \text{ (LENGTH} - 3.9651)^2]^{1/2} \quad .$$

A 95 percent confidence interval on average LPROD is

$$\hat{\text{LPROD}} \pm t(0.975;\ 61) \, \hat{\sigma} \, (\hat{\text{LPROD}}) = \hat{\text{LPROD}} \pm 2.000 \, \hat{\sigma} \, (\hat{\text{LPROD}}) \quad .$$

We calculate predicted mean productions and 95 percent confidence intervals for various values of length by exponentiating.

LENGTH	$\hat{\text{PROD}}$	LWR 95 PCT CONF BND	UPR 95 PCT CONF BND
3.00	35.02	27.67	44.33
3.50	59.68	51.98	68.54
4.00	101.72	92.92	111.34
4.50	173.35	149.12	201.52
4.75	226.30	185.53	276.03

The upper confidence bounds are about 50 to 60 percent greater than the lower confidence bounds at the extremes of the range (i.e. 3.00 and 4.75 mm) and about 20 percent greater in the middle of the range (i.e. 4.00 mm). Such precision (or lack of precision) may be adequate for assessing general trends in production with increasing length, but is probably not adequate for using lengths as a surrogate response for production.

F. SUMMARY AND CONCLUSIONS

We have studied the relationship between production and length in three data sets. We saw that the extent of association varied considerably in the different tests. This suggests that no generalization can be made about the association between these variables across tests. In some tests they will be more strongly associated than in others.

Both prediction and length responses pertained to beakers in the LeBlanc data sets but pertained to individual daphnids in the Chapman data set. The degree of association between production and length was much stronger in the Chapman data than in the LeBlanc data. This suggests that responses should be collected on a per daphnid basis rather than on a per beaker basis if the two variables are to be associated by regression models.

Just 21 day production and 21 day lengths were used in the regression models fitted in this section. Perhaps better association could be attained if increments of production were related to intermediate values of length and changes in these values. Such an effort, involving perhaps 7, 14 and 21 day lengths and production, would be somewhat more complex than working just with 21 day responses.

FILE NONAME (CREATION DATE = 07/09/81)
SUBFILE TESTA

SCATTERGRAM OF (DOWN) FERT21
(ACROSS) LENGTH21

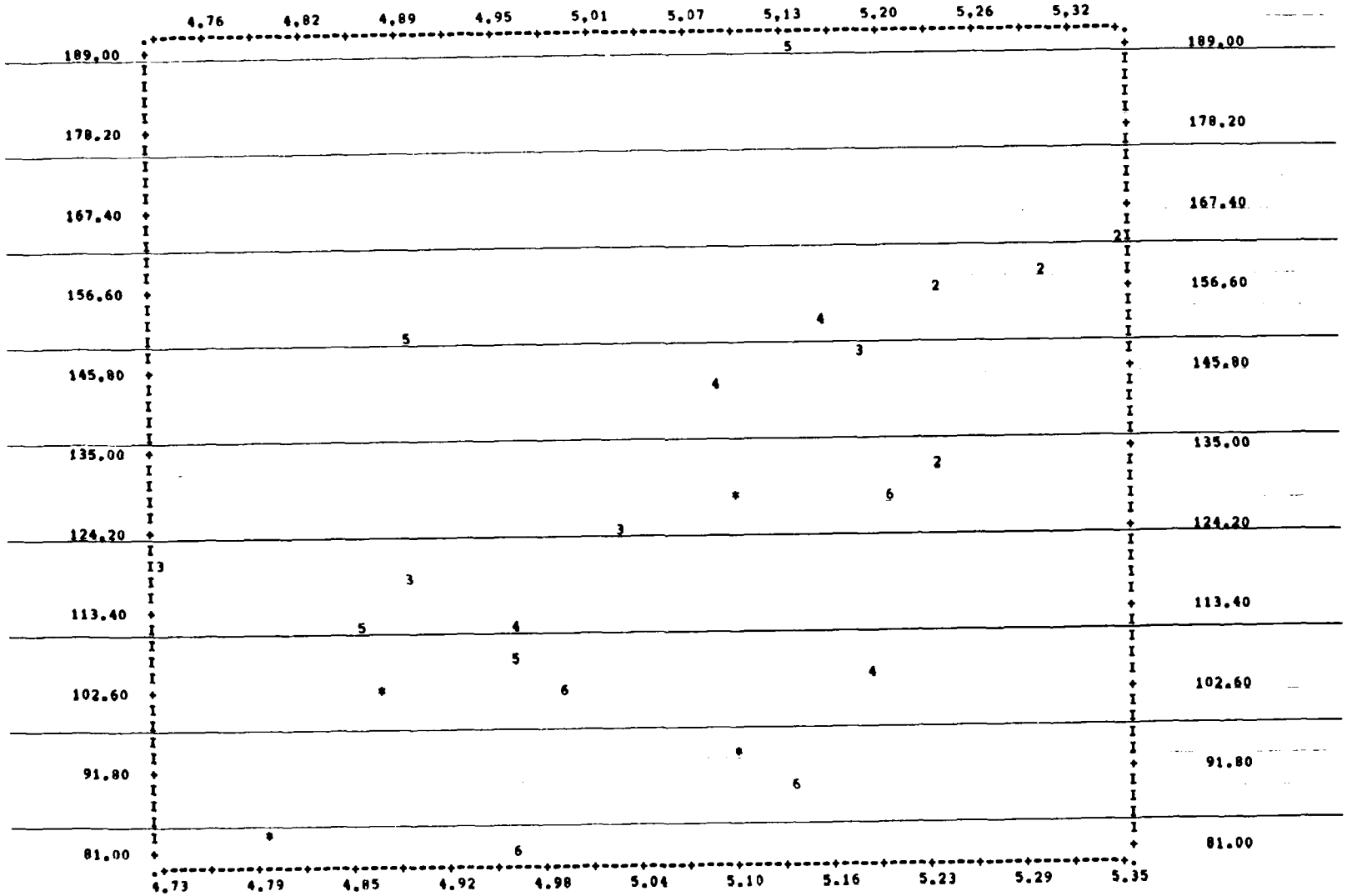


Figure VIII.1. 21 day cumulative reproduction per surviving adult per beaker versus 21 day average length per beaker - LeBlanc Test A.

FILE NONAME (CREATION DATE = 07/09/81)
SUBFILE TESTB

SCATTERGRAM OF (DOWN) FERT21
(ACROSS) LENGTH21

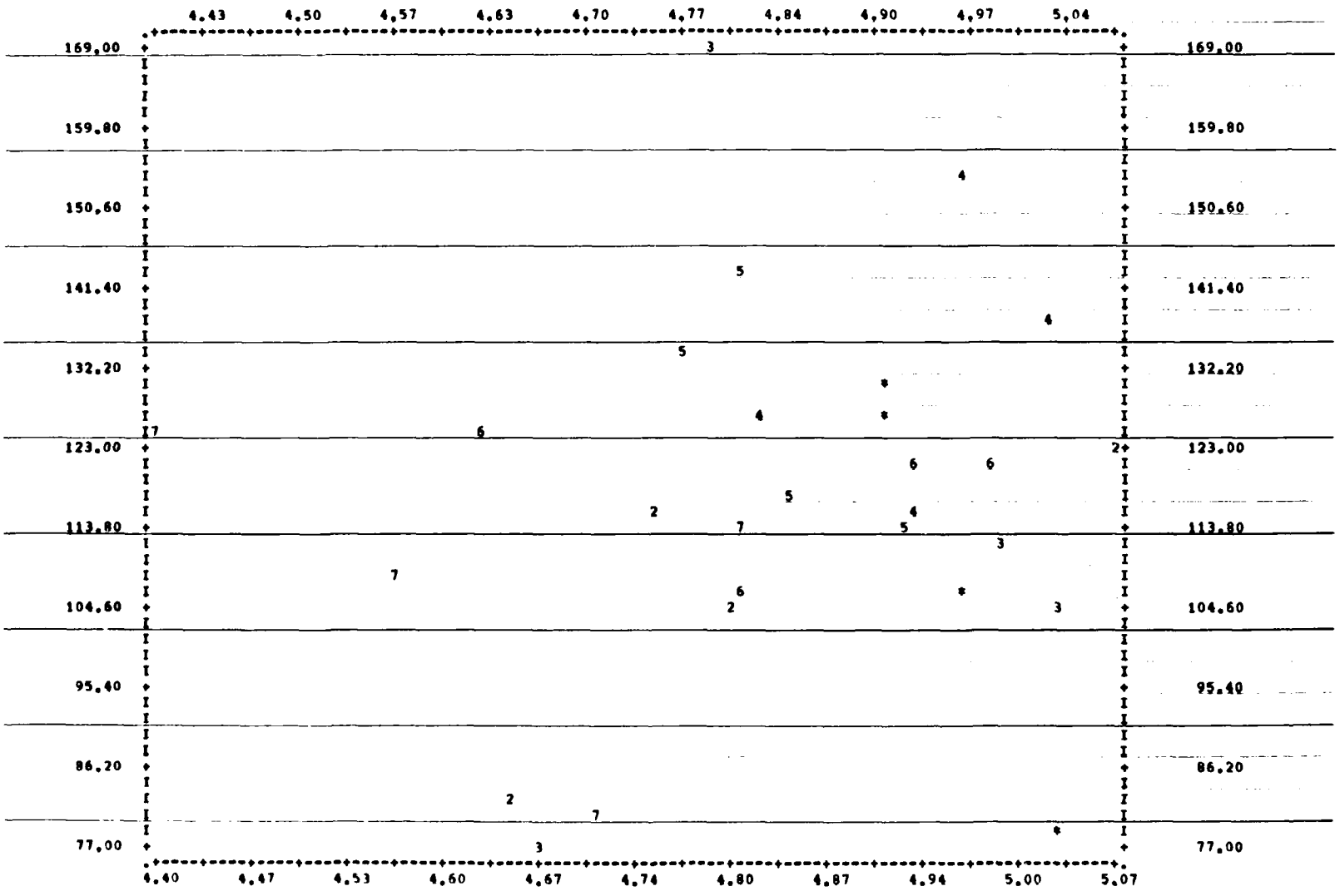


Figure VIII.2. 21 day cumulative reproduction per surviving adult per beaker versus 21 day average length per beaker - LeBlanc Test B.

FILE NONAME (CREATION DATE = 07/08/81)

SCATTERGRAM OF (DOWN) TOTALYNG
(ACROSS) LENGTH

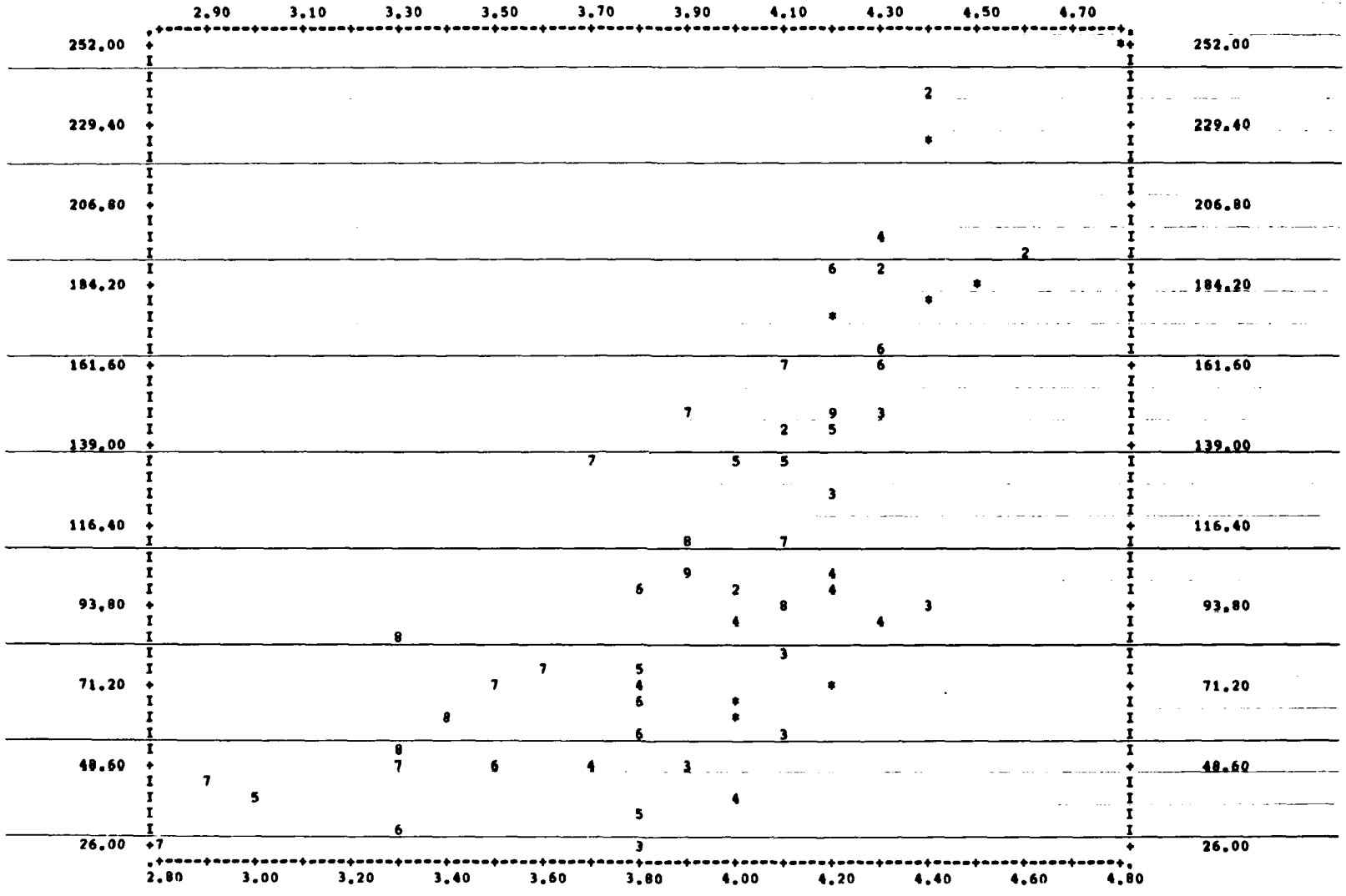


Figure VIII.3. 21 day cumulative reproduction versus 21 day length for surviving adults - Chapman Beryllium.

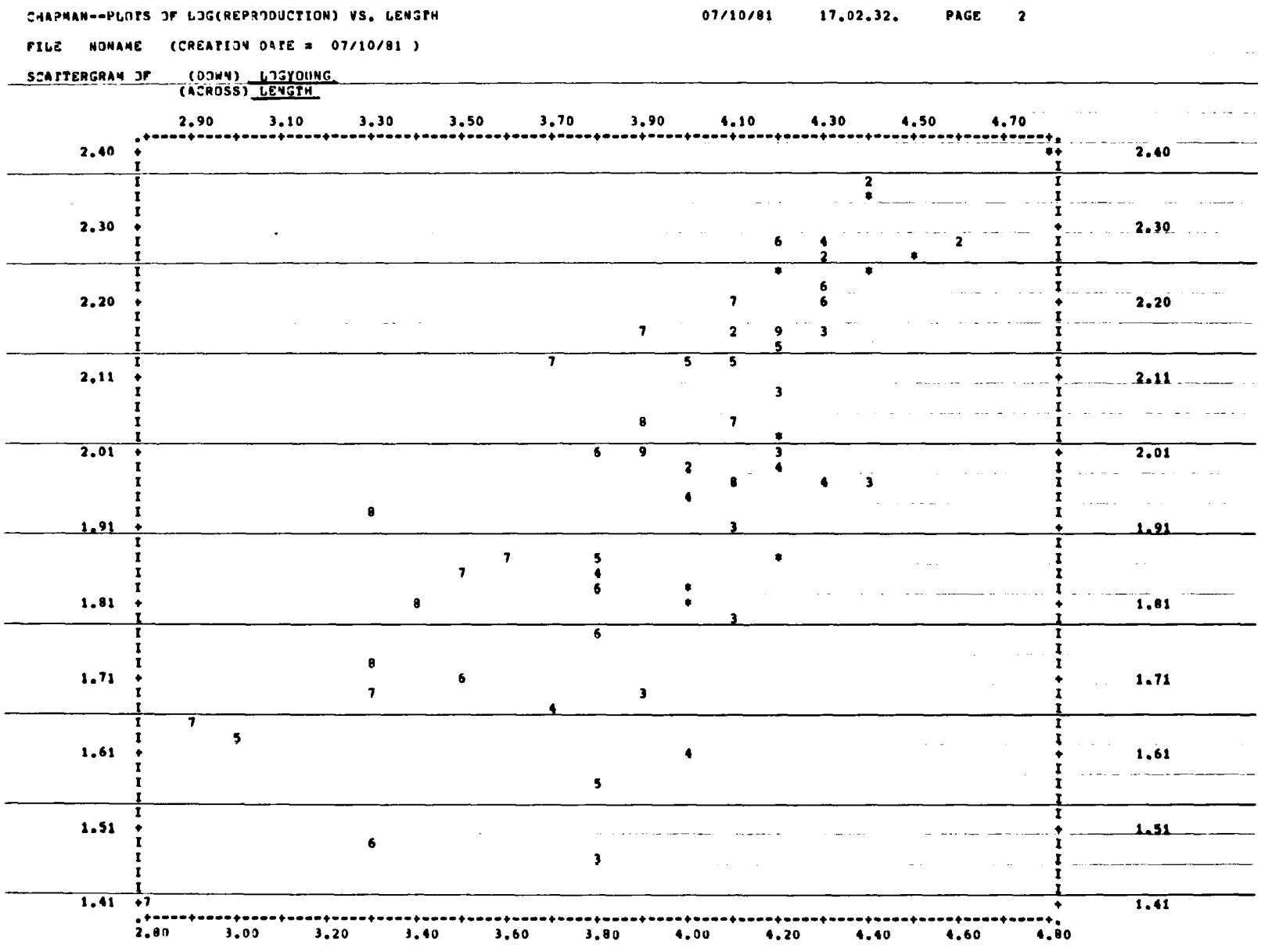


Figure VIII.4. 21 day $\log_{10}$ (cumulative) reproduction versus 21 day length for surviving adults - Chapman Beryllium

REGRESSION TITLE. . . . . .BIONOMICS TEST A, REPRODUCTION VS 21-DAY LENGTHS
DEPENDENT VARIABLE. . . . . 3 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R .5365 STD. ERROR OF EST. 24.3219
MULTIPLE R-SQUARE .2878

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	5259.793	1	5259.793	8.891	.00688
RESIDUAL	13014.166	22	591.553		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT		-338.73666					
LENGTH21	5	91.87210	30.810	.536	2.982	.007	1.000000

Figure VIII.5. Simple linear regression of 21 day cumulative reproduction per surviving adult versus average 21 day length - LeBlanc Test A.

REGRESSION TITLE. . . . . .BIONOMICS TEST B, REPRODUCTION VS 21-DAY LENGTHS
DEPENDENT VARIABLE. . . . . 3 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R .1258 STD. ERROR OF EST. 21.5606
MULTIPLE R-SQUARE .0158

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	194.320	1	194.320	.418	.52359
RESIDUAL	12086.394	26	464.861		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT		35.59855					
LENGTH21	5	16.98709	26.274	.126	.647	.524	1.000000

Figure VIII.6. Simple linear regression of 21 day cumulative reproduction per surviving adult versus average 21 day length - LeBlanc Test B.

REGRESSION TITLE. . . . . .CHAPMAN--REGRESSION OF REPRODUCTION VS LENGTH
DEPENDENT VARIABLE. . . . . 13 NYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R .7431 STD. ERROR OF EST. 38.4613
MULTIPLE R-SQUARE .5522

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	111284.424	1	111284.424	75.229	.00000
RESIDUAL	90235.513	61	1479.271		

VARIABLE	COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT	-304.22086					
LENGTH 12	105.15978	12.124	.743	8.673	-.000	1.000000

Figure VIII.7. Simple linear regression of 21 day cumulative reproduction versus 21 day length for surviving adults - Chapman Beryllium.

REGRESSION TITLE. . . . . .CHAPMAN--REGRESSION OF LOG(REPRODUCTION) VS LENGTH
DEPENDENT VARIABLE. . . . . 14 LOGYOUNG
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R .7713 STD. ERROR OF EST. .1552
MULTIPLE R-SQUARE .5949

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	2.158	1	2.158	89.579	.00000
RESIDUAL	1.469	61	.024		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT		.15515					
LENGTH	12	.46306	.049	.771	9.465	-.000	1.000000

Figure VIII.8. Simple linear regression of $\log_{10}$ (21 day cumulative reproduction) versus 21 day length for surviving adults - Chapman Beryllium.

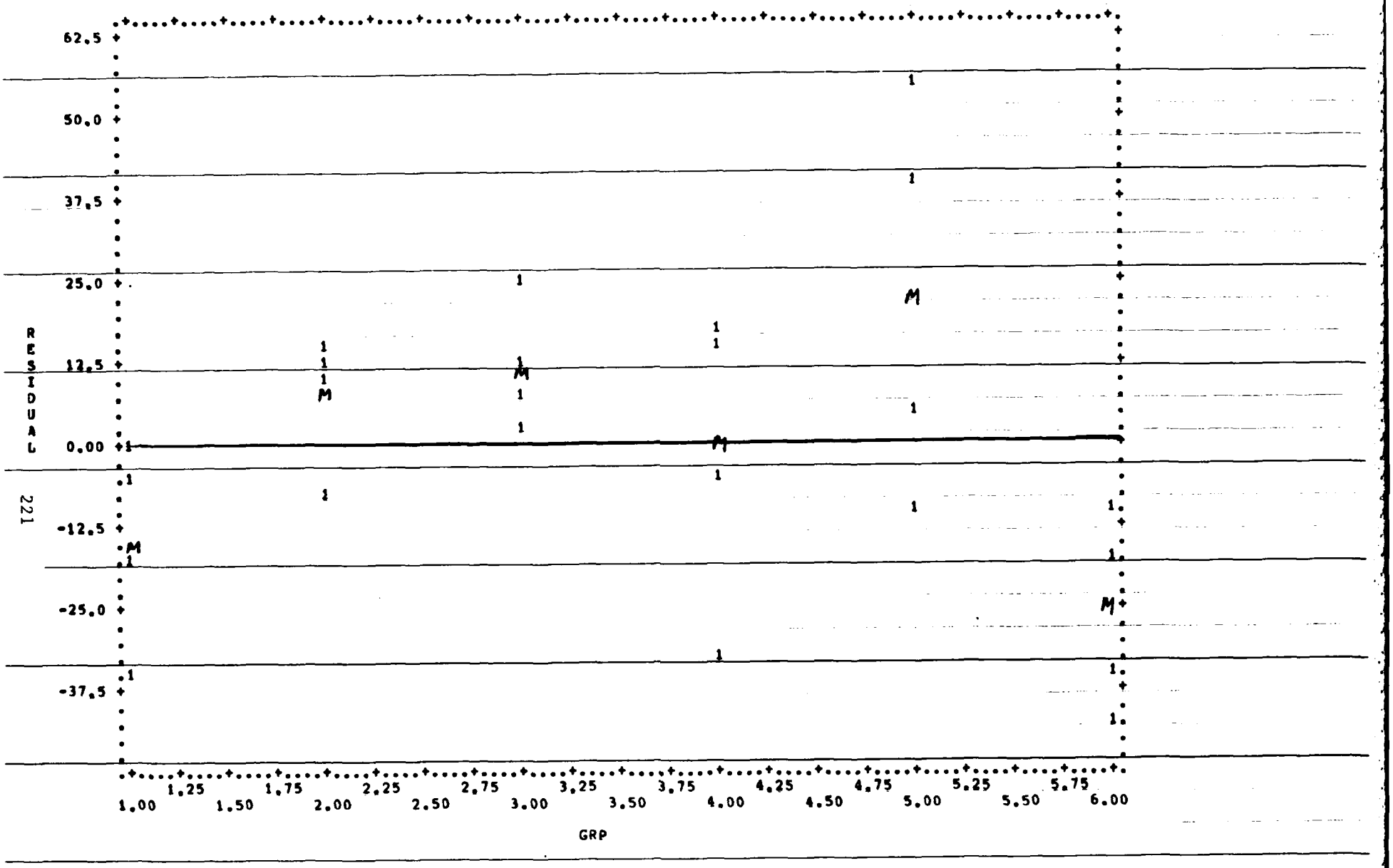


Figure VIII.9. Residuals from simple linear regression of reproduction on length plotted versus group number - LeBlanc Test A (Group 7 responses deleted) (M corresponds to group average of the residuals).

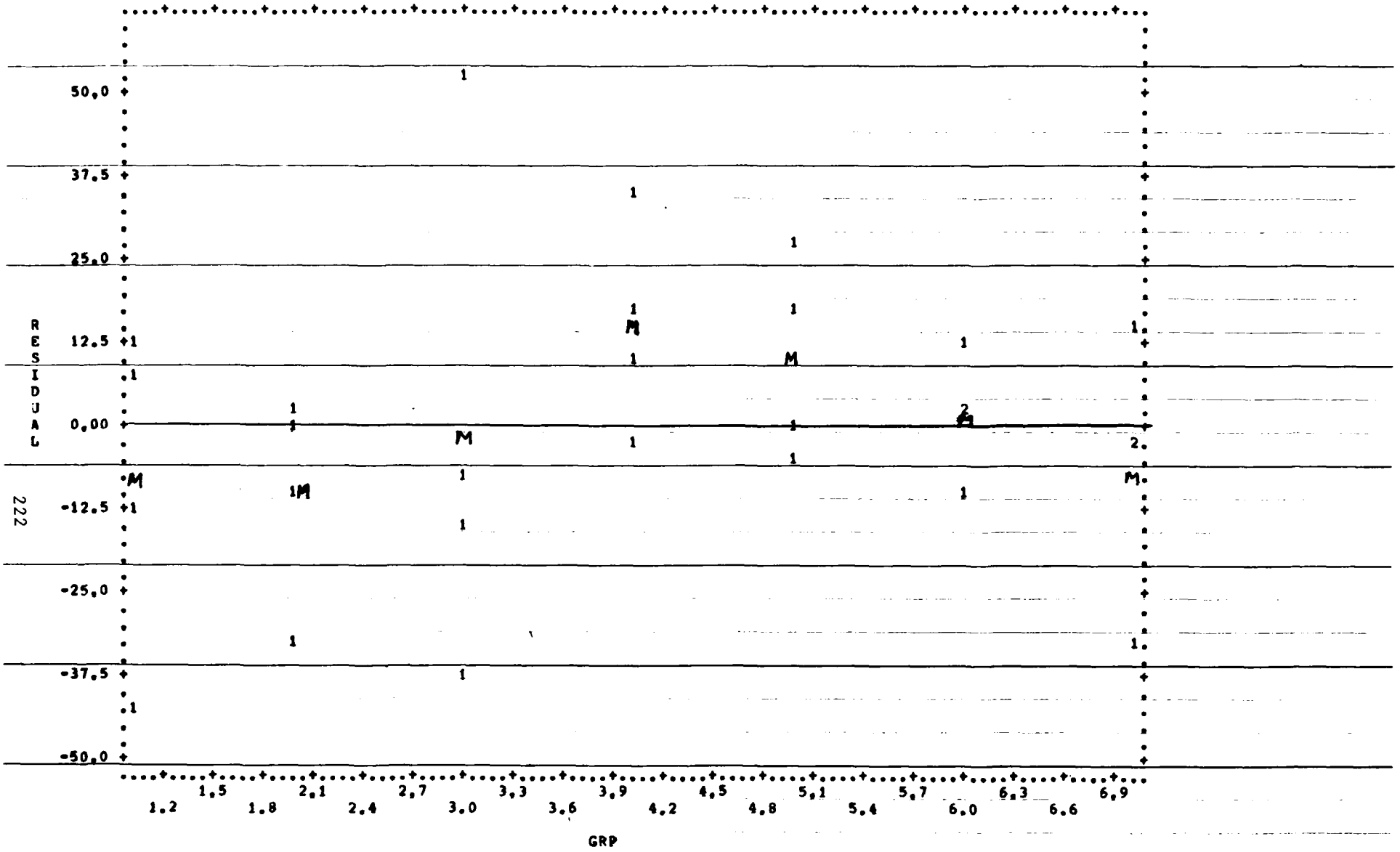


Figure VIII.10. Residuals from simple linear regression of reproduction on length plotted versus group number - LeBlanc Test B (M corresponds to group average of the residuals).

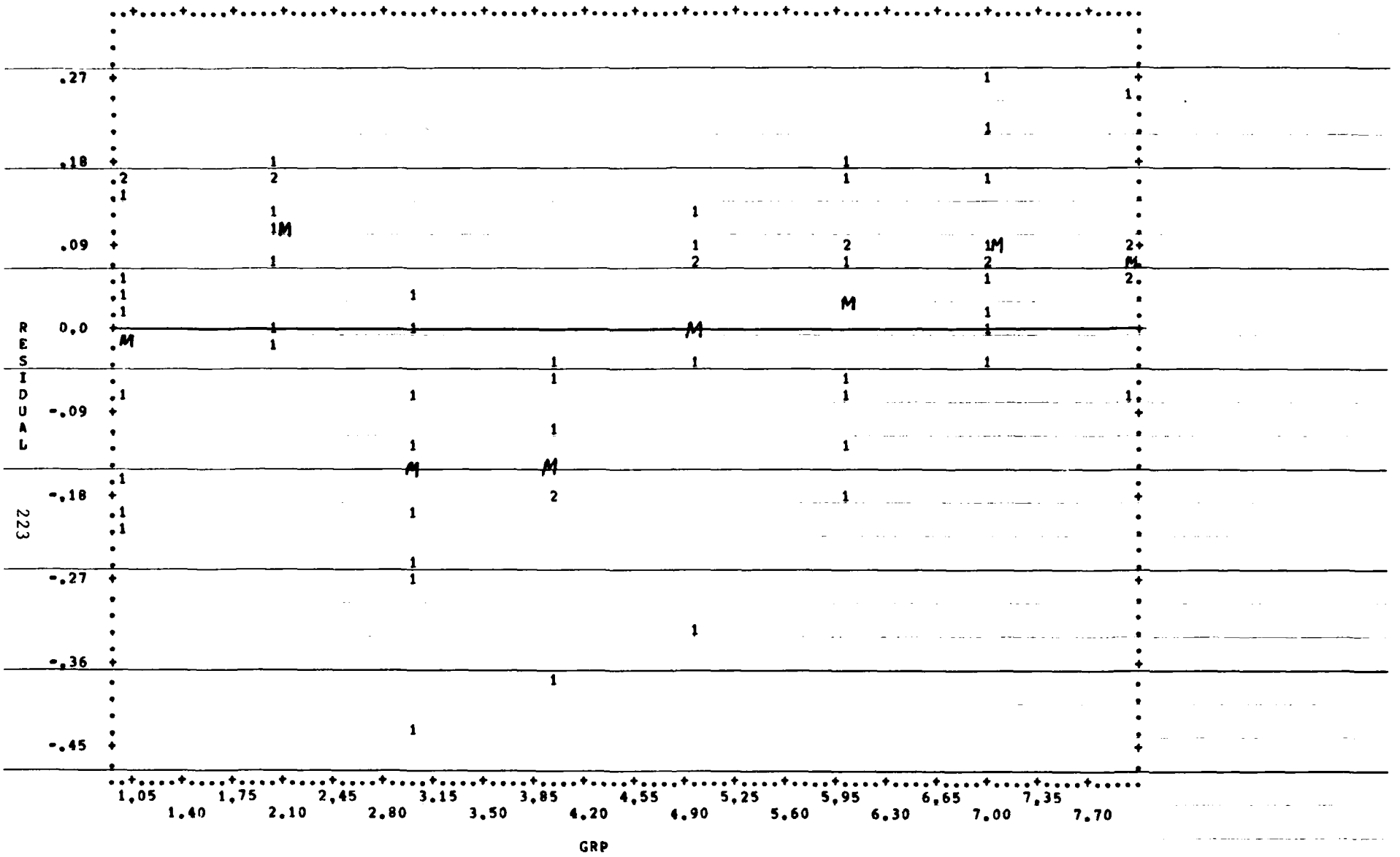


Figure VIII.11. Residuals from simple linear regression of $\log_{10}$ (reproduction) on length plotted versus group number - Chapman Beryllium (M corresponds to group average of the residuals).

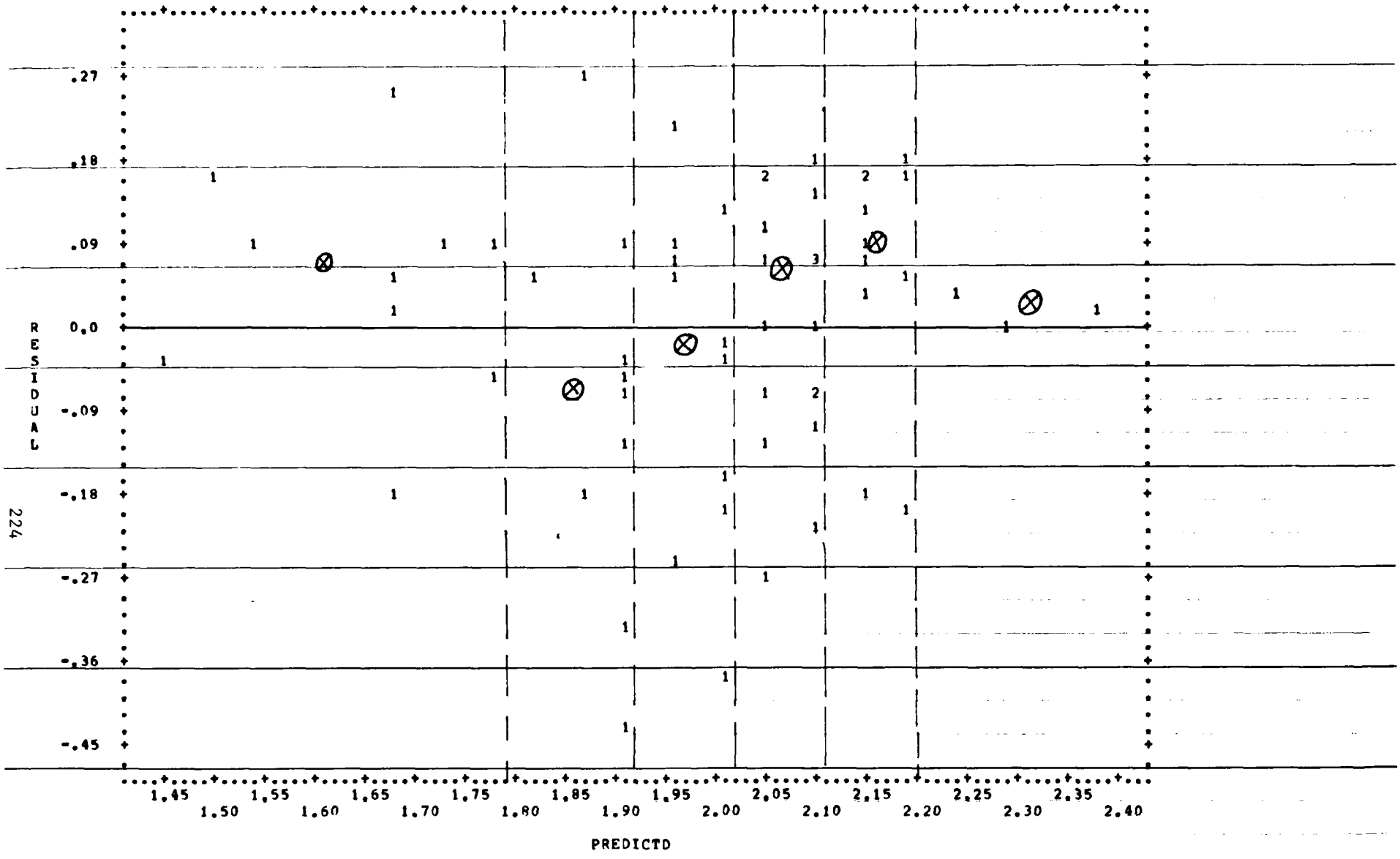


Figure VIII.12. Residuals from simple linear regression of $\log_{10}$ (reproduction) on length plotted versus predicted - Chapman Beryllium.

IX. TESTING FOR CONCENTRATION RELATED EFFECTS ON MORTALITY

A. INTRODUCTION

After we have carried out preliminary graphical displays, tests for beaker to beaker heterogeneity within groups, outlier detection procedures, and adjustments to account for beaker to beaker heterogeneity we are ready to proceed to the main portion of the data analysis. This involves comparing responses across treatment groups to arrive at interferences about what constitutes an "acceptable" concentration. In this and subsequent sections we discuss a number of hypothesis testing and estimation procedures to compare the responses obtained in the treatment groups with those in the control group(s). Comparisons are made for mortality, length and reproduction responses. We first discuss comparisons of mortality rates across groups.

This section discusses a number of hypothesis testing procedures to make inferences about which treatment groups are statistically significantly different from the control group(s). The chi square test for homogeneity, the measure of association tests based on various Goodman and Kruskal measures, the Cochran Armitage test and extensions, and Williams' test are discussed and illustrated. The chi square test is an overall, shotgun type test while the other tests are one sided tests, tailor made to be sensitive to monotone alternatives. They would thus be expected to be more sensitive than the chi square test to the types of alternative hypothesis to be expected in aquatic toxicity data.

The following sections discuss dose response curve estimation based on probit models, inferences based on these fits, and confidence interval inferences to compare treatment group and control group responses, based either on unadjusted response rates or on response rates smoothed across groups by fitting dose response curves.

As discussed in Feder and Collins [1], inferences based on hypothesis testing procedures have several severe drawbacks. First of all they are based on the notion of "statistical significance" and do not account at all for "biological significance". The notion of statistical significance is dependent on the sample size and on the variability of the responses as well as on the magnitudes of the effects observed in the test. Thus effects of considerable biological importance could be declared not statistically significant if the sample sizes are too small or the variability of responses is too great. Conversely, biologically trivial effects could be strongly statistically significant if the sample sizes are very large.

Inferences based on dose response curves can take biological significance into account by formulating the statistical problem as determining that toxicant concentration that results in a 10 percent or a 25 percent, etc. increase in mortality over the control group rate or a decrease of 0.5 mm or 1 mm, etc in average length as compared to the control groups, or a decrease in production of 25 or 50, etc offspring per adult as compared with the control group.

Furthermore the effects of small sample sizes or variable responses on estimates of "safe concentrations" are opposite depending on whether tests of hypothesis or dose response curves are used as the basis of making inferences. Namely if the "safe concentration" is defined as the highest concentration whose response rate is not statistically significantly different than the control group rate, then small sample sizes or variable responses will result in not rejecting H_0 for moderate differences between treatment and control group responses. This, in turn, will produce an increase in the reported "safe concentration". The smaller and less precise the toxicity test is, the greater will be the reported "safe concentration". By contrast, if the "safe concentration" is defined as the lower confidence bound on that concentration which produces a given increase, for example 10 percent, in mortality above the control group rate or a given decrease in length or production as compared with the control group, then small sample sizes or variable responses will result in longer confidence intervals and therefore reduced lower bounds. This, in turn, will produce a decrease in the reported "safe concentration".

Thus inferences based on the percentiles of dose response curves yield more conservative estimates from toxicity tests with limited information than from toxicity tests with ample information. Inferences based on tests of hypotheses yield less conservation estimates from toxicity tests with limited information than from toxicity tests with ample information.

Opinion: Inferences based on estimated percentiles of dose response curves are more appropriate than those based on hypothesis tests because they explicitly incorporate the notion of biological significance into the reported value and because they yield more conservative estimates from toxicity tests that provide limited information.

B. ADJUSTMENT OF DATA

Section V contains adjustment procedures to account for the presence of beaker to beaker heterogeneity in mortality and length responses. Mortality responses were adjusted by reducing the actual numbers of responses and daphnids per beaker to effective responses and sample sizes. Effective degrees of freedom were also calculated. Inference procedures on lengths were adjusted for beaker to beaker heterogeneity by carrying out inferences on a per beaker basis but augmenting the degrees of freedom associated with error estimates by utilizing information about the extent of variation in responses among beakers within groups in relation to the variation among daphnids within beakers.

In this section we illustrate the use of various hypothesis testing procedures on the mortality data from LeBlanc's Tests A and B and Goulden's Isophorone Test. The adjusted sample sizes, numbers of responses, and degrees of freedom for LeBlanc's Tests A and B are displayed in Tables V.3 and V.4, respectively. Goulden's isophorone mortality data show no evidence of heterogeneity of responses within groups among the beakers with multiple daphnids. Thus no adjustments are needed.

C. SEQUENTIAL TESTS

The classical approach to carrying out overall tests of hypothesis is to include all groups in the test statistic. If the test fails to reject then conclude that there is no statistical evidence of concentration related increases in mortality. If the test rejects then follow up with a multiple comparisons procedure such as Dunnett's or Williams' procedures to determine which treatment groups differ significantly from the control group and thereby arrive at an MATC. This two part procedure can be combined into a single procedure utilizing a sequential testing approach. This approach can be applied for any test procedure. The test is first carried out using all the treatment groups. If the test rejects the null hypothesis, the highest treatment group is deleted and the test is repeated on the remaining groups. No simultaneity adjustments are made. Each time the test rejects, another group is deleted. This process is continued until the test no longer rejects. The highest remaining treatment group is declared to be the MATC. If the mortality rate is a monotone increasing function of concentration, then any group for which the mortality rate is the same as the control rate has probability at most α (where α is the type 1 error level) of being declared significantly different from the control group. This holds simultaneously for all such groups. This is because any group with a mortality rate the same as the control rate can be declared significantly different from the control group only if the highest concentration group among those not different from the control group is; this has probability at most α. The idea of carrying out tests in a sequential manner was communicated to me by Dr. David Schoenfield.

D. CHI SQUARE TEST OF HOMOGENEITY ACROSS TREATMENT GROUPS

The most commonly used, although not the most appropriate, overall test for differences in mortality rates across treatment and control groups is the chi square test for homogeneity. It is analagous to the "shotgun" analysis of variance F test for quantitative responses. It is not entirely appropriate for testing homogeneity of mortality responses in aquatic toxicity tests because the treatment groups have a natural ordering (i.e. concentration level), the anticipated alternative may be of a particular type (e.g. increased mortality), and the magnitude of response may be monotone increasing or decreasing with increasing group number (e.g. increasing mortality or decreasing average production with increasing concentration). The chi square test is not designed to take any of this structure into account. It is thus relatively inefficient compared to other test procedures that are designed to be sensitive to such one sided, monotone alternatives. Several such alternative tests are discussed in later subsections.

The form of the chi square test is well known and is discussed in a number of books, papers, and reports. Feder and Collins [1], Subsection XI.B present expressions for the test statistic. Standard textbooks such as Dixon and Massey [8], pp. 240-243 or Freund [9], pp. 287-290 discuss the chi square test in detail and illustrate its application. The test is implemented as a standard feature in most statistical computing systems. For example, the procedure PROC FREQ in the SAS system (Barr et al [10]) or the program BMDP1F in the BMDP system (Dixon and Brown, [7]), can be used

to carry out this test. The EXAX2 program (Feder and Willavize, [2]) will also carry out the chi square test, using asymptotic or exact, small sample theory depending on the magnitudes of the expected cell frequencies.

We present below applications of the chi square test of homogeneity to the mortality responses in LeBlanc's Tests A and B and in Goulden's Isophorone Test. In LeBlanc's Test A we carry out separate comparisons, using the water control and using the solvent control groups as standards. We carry out separate comparisons against the two control groups rather than combining them because we demonstrated in Subsection VII.B significant differences in their mortality rates.

D1. LeBlanc Test A - Comparison With Water Control Group

We adjust for the effects of beaker to beaker heterogeneity within groups by reducing the actual sample sizes and numbers of responses to effective sample sizes and numbers of responses, as shown in Table V.3. Group 6 is discounted more than the others due to the widely disparate mortality rates among the four beakers. We pool the adjusted sample sizes across beakers within groups and carry out the usual chi square test as if there was no beaker to beaker heterogeneity. The estimated degrees of freedom for each group is given in Table V.3 and so the pooled degrees of freedom is 30 x 5 + 3 = 153. We should compare the "chi square" statistic to the upper 95 percent point of five times an F-distribution with degrees of freedom 5 and 153 rather than to a chi square distribution with 5 degrees of freedom. Since the difference between the percentiles corresponding to 153 d.f. and infinite d.f. is so minute, we ignore this adjustment.

Program BMDP1F would not accept the nonintegral "sample sizes" and numbers of "responses" that result from the adjustment process. It truncates all frequencies down to the next lowest integers. This program could therefore not be used to compare adjusted frequencies across groups. It should be noted that there is no theoretical reason for the program to carry out such truncations.

The EXAX2 program does allow the use of noninteger "frequencies" and so we used this program for the examples below. Figure IX.1 contains the results of the chi square test of homogeneity across groups for the LeBlanc Test A 21 day mortality responses. The test is based on the adjusted responses and sample sizes, as discussed previously. The water control group is used for comparison purposes. Although the expected adjusted frequencies in group 6 are less than 5.0, we base the test on asymptotic chi square theory since most of the cell frequencies are rather large and those in group 6 are moderate.

The observed chi square value of 177.63 (with 5 d.f.) is very highly significant. It is quite clear that the response rate in group 7 differs from those in the other groups. Perhaps the response rate in group 6 does also.

We delete group 7 from the data and recalculate the test statistic. Figure IX.2 contains the results of this test. The observed chi square value is reduced to 9.47. The test, based on asymptotic theory, is

marginally significant ($\alpha = 0.05$ or $\alpha = 0.06$ depending on whether the chi square distribution with 4 d.f. or the F-distribution with 4 and 123 d.f. is used for comparison). There is thus borderline statistical evidence of differences in response rates among the groups. Based on the appearances of Figures IX.2 and II.1, it is clear that the response rate in group 6 differs from those in the other groups.

The procedure could be continued by deleting group 6 and continuing. However, based on the appearance of Figure II.1 and the significance level in Figure IX.2, it is clear that the resulting chi square test would be nonsignificant. The process was thus stopped at this point and group 5 was declared to be the MATC group.

D2. LeBlanc Test A - Comparison With Solvent Control Group

The data and the adjustments are the same as those discussed in paragraph D1 except that the solvent control group (group 2) is used for comparisons in place of the water control group (group 1). The tests are again carried out with the EXAX2 program. The results are similar to those obtained by using the water control group.

Figure IX.3 contains the results of the chi square test of homogeneity across groups for the LeBlanc Test A 21 day mortality responses, using the solvent control group. The test is based on asymptotic theory. The observed chi square value of 199.14 (with 5 d.f.) is very highly significant. The response rate in group 7 differs from those in the other groups.

We recalculate the test statistic after deleting group 7 from the data. Figure IX.4 contains the results of this test. The observed chi square is reduced to 14.14. This is still very highly significant ($\alpha = 0.007$). Thus there is still strong statistical evidence of differences in response rates among the groups. Based on Figures IX.4 and II.1, it appears as if the response rate in group 6 differs from those in the other groups.

We recalculate the test statistic once more after deleting both groups 6 and 7 from the data. Figure IX.5 contains the result of this test. The observed chi square is now 2.99 with 3 d.f. which is not statistically significant ($\alpha = 0.39$). The process is thus stopped at this point and group 5 is again declared to be the MATC group.

D3. LeBlanc Test B - Comparison With Combined Water and Solvent Control Groups

The adjusted sample sizes and responses are shown in Table V.4. The adjustment is very similar to that for the Test A data. Group 7 is discounted more than the others due to the widely disparate mortality rates among the four beakers. Since there was no statistically significant differences between the mortality rates in the water and solvent control groups, these two groups were combined into a common control group for the purpose of comparison with the treatment group responses. Other than that, the tests were carried out in the same manner as those for LeBlanc's Test A data. In particular the tests were carried out using EXAX2 and asymptotic analyses.

Figure IX.6 contains the results of the chi square test of homogeneity utilizing all the treatment groups and the combined control groups. The observed chi square value of 27.13 (with 5 d.f.) is very highly significant. The response rate of group 7 differs from those in the other groups.

We delete group 7 from the data and recalculate the test statistic. Figure IX.7 contains the results of this test. The observed chi square value is now 5.82 with 4 d.f. which is not statistically significant ($\alpha = 0.21$). The process is stopped at this point and group 6 is declared to be the MATC group.

D4. Goulden Isophorone Test

Attention is confined to the three beakers per group containing multiple daphnids (5 daphnids per beaker). It was shown in Subsection III.B that there is no statistical evidence of heterogeneity among beakers within groups for this test. Thus we do not adjust the sample sizes and numbers of responses prior to pooling data across beakers within groups and carrying out comparisons of response rates across groups. As there is just one control group (group 1), there is no issue about the comparability of results in water and solvent control groups. Other than those considerations, the tests were carried out in the same manner as those for LeBlanc's data from Tests A and B. In particular, the tests were carried out using EXAX2 and asymptotic analyses.

Figure IX.8 contains the results of the chi square test of homogeneity utilizing all the treatment groups and the control group. The observed chi square value of 36.50 (with 5 d.f.) is very highly significant. The mortality rate in group 6 is considerably higher than those in the other groups.

We delete group 6 from the data and recalculate the test statistic. Figure IX.9 contains the results of this test. The observed chi square has been reduced to 14.17 (with 4 d.f.), which is still highly significant ($\alpha = 0.007$). The mortality rate in group 5 appears to differ from those in the other groups.

We delete group 5 from the data and recalculate the test statistic. Figure IX.10 contains the results of this test. The observed chi square is now 1.78 with 3 d.f., which is not statistically significant ($\alpha = 0.62$). The process is stopped at this point and group 4 is declared to be the MATC group.

E. ONE SIDED, MEASURE OF ASSOCIATION TESTS FOR ORDERED CONTINGENCY TABLES

The shotgun chi square test, although the most commonly used test of homogeneity of response rates, is not the most appropriate test for application to aquatic toxicity data. The reasons for this were discussed at the beginning of Subsection D, above. Tests of hypothesis that are designed to detect one sided, monotone alternatives are more sensitive to and thus more appropriate for the kinds of alternatives relevant in aquatic toxicity

tests. One approach to the construction of one sided tests is by means of measures of association for ordered contingency tables. Goodman and Kruskal [11,12] have derived and reported on a number of measures. Feder and Collins [1], Subsection XI.C, discuss a number of these measures and inferences based on them in some detail. Measures discussed there include Goodman and Kruskal's gamma, Kentall's τ_b, Stuart's τ_c, and Somer's d.

These measures can be thought of as ordered contingency table analogs of correlation coefficients for quantitive responses. However for a given table each of these measures take on different values and so it is difficult to ascribe physical meaning to any of the values. Thus we do not recommend using the values of these measures as indicators of the strength of a toxicant. However for each of the measures a value of zero means no monotone association between group number and mortality rate. Positive or negative values of the measures mean positive or negative associations, respectively. It should be noted that, just as with correlation coefficients, a measure of monotone association can be zero in the presence of a strong but nonmonotone association. Thus these measures of association can be used to test null hypotheses of homogeneity of mortality rates across groups against alternatives of one sided, monotone trends. Procter [13] has shown that such tests are much more powerful against one sided, monotone alternatives than is the shotgun, chi square test.

In order to use estimates of measures of association for statistical inferences, it is necessary to know something about their distribution, in particular their variability around the population value. Goodman and Kruskal [12] derive asymptotic (normal) distributions of these estimates by means of the delta method and present asymptotic standard errors. Brown and Benedetti [14] calculate improved standard error estimates for the various measures, that are more appropriate studentizing factors for testing the null hypothesis that those measures are zero. They show empirically that their standard error estimates yield better approximations to the nominal type 1 errors in small and moderate samples than do the Goodman and Kruskal standard error estimates. Furthermore, a very interesting attribute of the Brown and Benedetti standard error estimates is that even though each of the measures of association in general have different numerical values, the "t ratios" formed by normalizing the measures by their respective standard errors have identical values. Thus there is just one t-ratio associated with all five measures (γ, τ_b, τ_c, and two d's). This t-ratio can thus be interpreted without ambiguity.

Brown and Benedetti report, based on a simulation study, that for sample sizes in excess of 100, the "t-ratios" can be treated as normal random variables for the purpose of testing hypotheses about the significance of the relation. For sample sizes, N, less than 50 they recommend comparing the t-values to a t-distribution with approximate degrees of freedom 0.4N. See the Brown and Benedetti paper for further details.

Feder and Collins empirically illustrate the increased sensitivity to monotone alternatives, of the measure of association test relative to the chi square test by applying both tests to several sets of artificial data constructed to reflect mild, moderate, and strong trends. In each case the measure of association test is much more highly significant. See Feder and Collins for details.

The measure of association tests have been incorporated in the BMDP program, BMDP1F [7]. As remarked in the previous subsection, this program will not accept nonintegral values for sample sizes and numbers of responses. Thus this program cannot be used directly on the adjusted data values. An indirect method of adjusting the test procedure for the presence of beaker to beaker heterogeneity is to carry out the test on the original, unadjusted values and then modify the estimated standard error upward and the t-ratio downward by a factor reflecting the heterogeneity adjustments. Degrees of freedom would be based on the degrees of freedom arrived at in the adjustment process. Consider for example LeBlanc's Test A data. There were $7 \times 4 \times 20 = 560$ daphnids used in this test. The adjusted sample sizes, numbers of responses, and numbers of degrees of freedom are displayed in Table V.3. The adjusted number of daphnids is $6 \times 4 \times 14.8 + 4 \times 1.45 = 361$ and the adjusted number of degrees of freedom is $6 \times 30 + 3 = 183$. We thus carry out the measure of association test based on the unadjusted frequencies, inflate the estimated standard errors of the various measures by the factor $[560/361]^{1/2} = 1.25$ and reduce the calculated t-ratio by this same factor. The resulting value is compared to a normal distribution. This indirect adjustment procedure, while intuitively reasonable, has not been studied theoretically. Its theoretical properties are therefore unknown.

We illustrate the application of the one sided measure of association test on the mortality data from Goulden's test on isophorone. Only the responses from the beakers with multiple daphnids were used. Since there was no statistical evidence of beaker to beaker heterogeneity within groups, no adjustments were carried out. The responses were pooled across beakers and comparisons among groups were based on 15 daphnids per group. The test was carried out in a sequential manner, as discussed in Subsection C. The results are shown in Figures IX.11 - IX.13.

Figure IX.11 displays the results of the chi square and measure of association tests applied to all the data. Recall that the chi square test is a two sided test whereas the measure of association test is a one sided test. The chi square test statistic is compared to a chi square distribution with 5 d.f. and is seen to be very highly significant. The measure of association t value is compared to a t-distribution with $90 \times 0.4 = 36$ d.f. The value, 5.730, is significant at $\alpha = 0.0000$. Thus both tests show strong statistical evidence of a concentration effect.

We delete group 6 from the data and recalculate the test statistics. The results are shown in Figure IX.12. The chi square test statistic, with 4 d.f., is significant at $\alpha = 0.007$. The measure of association t-value is compared to a t distribution with d.f. $75 \times 0.4 = 30$. The value, 2.583, is significant at $\alpha = 0.007$. Thus both tests show strong statistical evidence of a concentration effect and at about the same alpha level.

We now delete group 5 from the data and recalculate the test statistics. The results are shown in Figure IX.13. The chi square statistic, with 3 d.f., is significant at $\alpha = 0.6195$. The measure of association t-value is compared to a t distribution with d.f. $60 \times 0.4 = 24$. The value, 0.472, is significant at $\alpha = 0.32$. Thus neither test shows any statistical evidence of a concentration effect. However the α-level for the one sided test is much smaller than that for the chi square test, probably due to the increased

mortality in group 4. The process is stopped at this point and group 4 is declared to be the MATC group. In this example the chi square and measure of association tests arrive at the same conclusion.

F. TREATMENT GROUP VERSUS CONTROL GROUP PAIRWISE COMPARISONS-- WILLIAMS' TEST

A common approach to pairwise comparisons between the control group and the treatment groups is with Dunnett's or Williams' procedures. Within each group the observed frequencies are adjusted for beaker to beaker heterogeneity and then pooled across beakers. The response rate is calculated based on the pooled data. For qualitative response rate data such as mortality rates, an arc sin variance stabilizing transformation is carried out on the response rate within each group and comparisons are based on these transformed values. Dunnett's and Williams' procedures are discussed in a number of references [15,16,17,18,19]. Feder and Collins [1] discuss these procedures in Section XII. Chew [19] briefly describes Williams' test and presents tables for its implementation.

Williams' procedure is to be preferred to Dunnett's procedure if the mortality rate is a monotone increasing function of concentration, as it takes account of this monotonicty and is thus more sensitive in detecting weak to moderate trends. Williams [17], Section 4, compares the power of his test (which he calls the $\bar{t}$ test) to that of a one sided t-test and to Dunnett's test. The distribution theory is sufficiently complex, that the power comparisons must be carried out by Monte Carlo methods when there are three or more treatment groups. However Williams concludes on page 113, based on the results of the Monte Carlo experiments, that "It is evident...the superiority of the $\bar{t}$ test over both the one sided t-test... and Dunnett's one-sided test becomes more marked as k (the number of treatment groups-P.F.) increases. There is no doubt that the $\bar{t}$ test should be used in preference to these two tests...".

We illustrate Williams' method with several examples based on analysis of the 21 day mortality data. Consider first the 21 day mortality data from LeBlanc's Test A and compare treatment group responses to those in the water control group. We wish to determine which treatment group exhibit significantly greater mortality rates than the water control group. We adjust the sample sizes for beaker to beaker heterogeneity within groups, as indicated in Table V.3. The degrees of freedom assumed is 30 x 5 + 3 = 153. The basic and transformed responses, pooled across beakers within groups are:

Group (i)	1(Control)	3	4	5	6	7
Sample Size (n_i)	59.2	59.2	59.2	59.2	5.8	59.2
Response Rate ($\hat{p}_i$)	0.15	0.125	0.086	0.088	0.50	0.988
2 Arc Sin $\sqrt{\hat{p}_i} \equiv X_i$	0.79	0.72	0.60	0.60	1.57	2.92

Since these estimates are not in monotone sequence they need to be adjusted.

$$\hat{M}_1 = \hat{M}_3 = \hat{M}_4 = \hat{M}_5 = \frac{0.79 + 0.72 + 0.60 + 0.60}{4} = 0.68$$

$$\hat{M}_6 = 1.57, \hat{M}_7 = 2.92 \quad .$$

We declare the group i response rate to be significantly different from the control rate if

$$\hat{M}_i - X_1 > \bar{t} \, (1/n_i + 1/n_1)^{1/2}$$

The factors $\bar{t}$ can be obtained from Williams' tables (e.g. [18], Tables 1 and 3) corresponding to $\alpha = 0.05$ or $\alpha = 0.01$ and to $\nu = 153$. This yardstick is based on the asymptotic approximation that the variance of 2 arc sin $\sqrt{p}$ is 1/n. For simplicity, we use the values of $\bar{t}$ appropriate for equireplicated treatment groups, even though the effective sample size in group 6 is somewhat smaller than those in the other groups. We use the cutoff points corresponding to $\nu = 120$ d.f. and choose the $\bar{t}$'s sequentially, corresponding to the number of treatment groups.

Group 7 versus Group 1: $k = 5$, $\hat{M}_7 - X_1 = 2.92 - 0.79 = 2.13$,

$\bar{t} \, (1/n_7 + 1/n_1) = 1.772 \, (2/59.2)^{1/2} = 0.326.$

Thus M_7 is significantly greater than X_1.

Group 6 versus Group 1: $k = 4$, $\hat{M}_6 - X_1 = 1.57 - 0.79 = 0.78$,

$\bar{t} \, (1/n_6 + 1/n_1) = 1.765 \, (1/59.2 + 1/5.8)^{1/2} =$

0.768. Thus M_6 is significanly greater than

X_1 (but just barely).

Group 5 versus Group 1: $k = 3$, $\hat{M}_5 - X_1 = 0.68 - 0.79 < 0$. Thus M_5

is not significantly greater than X_1.

We thus stop the process and declare group 5 to be the MATC group. This result agrees with that previously arrived at based on the chi square test for homogeneity.

We apply this same procedure to the 21 day mortality data from LeBlanc's Test A, but comparing to the solvent control group rather than the water control group. We again adjust the sample sizes based on the results in Table V.3. The basic and transformed responses, pooled across beakers within groups are:

Group (i)	2(Control)	3	4	5	6	7
Sample Size (n_i)	59.2	59.2	59.2	59.2	5.8	59.2
Response Rate ($\hat{p}_i$)	0.037	0.125	0.086	0.088	0.50	0.988
2 Arc Sin $\sqrt{\hat{p}_i} \equiv X_i$	0.388	0.72	0.60	0.60	1.57	2.92
$\hat{M}_i$	0.388	0.64	0.64	0.64	1.57	2.92
$\hat{M}_i - X_1$		0.25	0.25	0.25	1.18	2.53

Since the X_i's are not in monotone sequence they need to be adjusted to the M_i's. The comparisons and yardsticks used are directly analagous to those discussed previously.

Group 7 versus Group 2: $k = 5$, $\bar{t}\,(1/n_7 + 1/n_2)^{1/2} = 1.772\,(2/59.2)^{1/2} = 0.326 < \hat{M}_7 - X_1$. Thus M_7 is significantly greater than M_2.

Group 6 versus Group 2: $k = 4$, $\bar{t}\,(1/n_6 + 1/n_2)^{1/2} = 1.765\,(1/59.2 + 1/5.8)^{1/2} = 0.768 < \hat{M}_6 - X_1$. Thus M_6 is significantly greater than M_2.

Group 5 versus Group 2: $k = 3$, $\bar{t}\,(1/n_6 + 1/n_2)^{1/2} = 1.754\,(2/59.2)^{1/2} = 0.322 < \hat{M}_5 - X_1$. Thus M_5 is not significantly different from M_1.

We thus stop the process and declare group 5 to be the MATC group. This result agrees with that arrived at above based on comparisons with the water control group. It also agrees with the MATC arrived at based on the chi square test for homogeneity.

We now apply Williams' procedure to the 21 day mortality data from LeBlanc's Test B, based on comparisons to the combined solvent and water control groups. We adjust the sample sizes based on the results in Table V.4. The basic and transformed responses, pooled across beakers within groups are:

Group (i)	0(Control)	3	4	5	6	7
Sample Size (n_i)	115.2	57.6	57.6	57.6	57.6	8.96
Response Rate ($\hat{p}_i$)	0.13	0.049	0.063	0.163	0.099	0.626
2 Arc Sin $\sqrt{\hat{p}_i} \equiv X_i$	0.74	0.44	0.51	0.83	0.64	1.826
$\hat{M}_i$	0.608	0.608	0.608	0.735	0.735	1.826
$\hat{M}_i - X_0$		-0.132	-0.132	-0.005	-0.005	1.086

We use the $\bar{t}$ values appropriate for equireplicated treatment groups, even though the effective sample size in group 7 is smaller than those in the other groups. Since the control group sample size is at least twice that of any of the treatment groups, we utilize the adjustment suggested by Williams [18, Section 2] to account for increased control group replication. Namely let c denote the control group sample size, let r denote the average treatment group sample size, and let w = c/r. In our example c = 115.2, r = 47.87, w = 2.41. Williams recommends adjusting $\bar{t}$ downward to $\bar{t} - 10^{-2}\,\beta(1 - 1/w)$, where values of β are given in Table 1, corresponding to k and ν. We use ν = 120 d.f.

Group 7 versus Control: $k = 5$, $\bar{t}\,(1/n_7 + 1/n_0)^{1/2} =$

$1.743\,(1/8.96 + 1/115.2)^{1/2} = 0.605 < \hat{M}_7 = X_0$.

Thus M_7 is significantly greater than M_0.

Since all the other $\hat{M}_i$'s are less than X_0, we conclude that M_6 is not significanly different from M_0. We thus stop the process and declare group 6 to be the MATC group. This result agrees with that previously arrived at based on the chi square test for homogeneity.

As the final example in this set we apply Williams' procedure to the 21 day mortality from Goulden's Isophorone Test, using the data from the multiply housed daphnids. There was no evidence of beaker to beaker heterogeneity so we carry out comparisons based on the unadjusted sample sizes. The basic and transformed responses, pooled across beakers within groups are:

Group (i)	1(Control)	2	3	4	5	6
Sample Size (n_i)	15.0	15.0	15.0	15.0	15.0	15.0
Response Rate ($\hat{p}_i$)	0.133	0.067	0.067	0.200	0.533	0.867
2 Arc Sin $\sqrt{\hat{p}_i} \equiv X_i$	0.75	0.52	0.52	0.93	1.64	2.39
$\hat{M}_i$	0.60	0.60	0.60	0.93	1.64	2.39
$\hat{M}_i - X_1$	-0.15	-0.15	-0.15	0.18	0.89	1.64

We use the $\bar{t}$ values appropriate for equireplicated treatment groups and $\nu = 14 \times 6 = 84$ d.f.

Group 6 versus Control: $k = 5$, $\bar{t}\,(1/n_6 + 1/n_1)^{1/2} = 1.79\,(2/15)^{1/2} = 0.650 < \hat{M}_6 - X_1$. This M_6 is significantly greater than M_1.

Group 5 versus Control: $k = 4$, $\bar{t}\,(1/n_5 + 1/n_1)^{1/2} = 1.773\,(2/15)^{1/2} = 0.647 < \hat{M}_5 - X_1$. Thus M_5 is significantly greater than M_1.

Group 4 versus Control: $k = 3$, $\bar{t}\,(1/n_4 + 1/n_1)^{1/2} = 1.762\,(2/15)^{1/2} = 0.643 < \hat{M}_4 - X_1$. Thus M_4 is not significantly greater than M_1.

We thus stop the process and declare group 4 to be the MATC group. This result agrees with that previously arrived at based on the chi square test for homogeneity.

In summary in all the examples we have considered, the chi square test and Williams' test lead to the same conclusions about the MATC. This is because there is generally either low mortality or high mortality observed, with few groups falling in the middle of the dose response curve. In such cases there are no borderline situations and so most reasonable test procedures will arrive at the same conclusion. This situation of course, does not hold in general.

G. ONE SIDED COMPARISONS BASED ON THE COCHRAN-ARMITAGE TEST

In the previous subsections of this section we have considered several procedures to test for the presence of statistically significant differences between treatment group and control group mortality rates. In Subsection D we considered the overall, analysis of variance type Pearson chi square test. In Subsections E and F we considered one sided tests that are designed to be more sensitive to monotone alternatives. In this subsection we consider a generalization of the chi square test, due to Cochran and Armitage, that is also more sensitive to one sided, monotone alternatives.

The Cochran-Armitage test is appropriate when the experimental groups possess an intrinsic ordering, such as is the case in aquatic toxicity tests. A score is attached to each group, so that an ordered scale is created. These scores are treated as predictor variables and the null hypothesis of equal response probabilities across groups is tested against an alternative of some type of trend of response probabilities with increasing score. Let $X_0, X_1, X_2, \ldots, X_k$ denote the scores assigned to the control group and the k treatment groups. Let $p_0, p_1, \ldots, p_k$ denote the mortality probabilities in these groups. Then the hypothesis can be expressed as

$$H_0: \quad p_0 = p_1 = \ldots = p_k$$

versus

$$H_1: \quad p_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_r X^r$$

Cochran and Armitage chose the simplest type of trend, a straight line trend, to test against but this test can be extended to polynomials of higher order. Thus we will test H_0 against the alternative hypothesis

$$H_1: \quad p_i = \beta_0 + \beta_1 X_i$$

This test has been described and illustrated in a number of references. Snedecor and Cochran [20], Section 9.11, Steel and Torrie [21], Section 22.10, Fleiss [22], Section 9.2, Cochran [23], and Armitage [24] discuss this procedure in some detail. In essence the homogeneity chi square with k degrees of freedom is partitioned into a single degree of freedom component to test for a straight line trend and a residual component with k-1 degrees of freedom to test for departures from linearity. The residual component can be further decomposed, if desired, into quadratic components, cubic components, etc. Expressions for the decomposition are given by Fleiss [22], Section 9.2, Equations (9.17) - (9.26).

The Cochran-Armitage procedure can be carried out rather easily utilizing any linear regression computer program with capabilities of performing weighted least squares fits. Namely the regression model

$$p_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \bar{p}\bar{q}/n_i)$$

is fitted to the data. The $\hat{p}_i$'s are the observed mortality rates based on an effective sample size of n_i daphnids. The ε_i's are independent variables, representing the random variation with mean 0 and variance $\bar{p}\bar{q}/n_i$. The probability $\bar{p}$ is the common value of all the p_i's under the null hypothesis and $\bar{q} = 1-\bar{p}$. It is estimated by the total (effective) number of deaths among all the groups divided by the total (effective) number of daphnids in all the groups. Thus the weight, w_i, for the i-th group is $n_i/\bar{p}\bar{q}$.

This procedure was carried out using the BMDP polynomial regression computer program, BMDP5R. See Dixon and Brown [7] for a detailed description of this program. Cubic polynomials were fitted rather then straight lines, however the Cochran-Armitage test can be carried out based on this output. The tests were carried out sequentially.

The chief objection to the Cochran-Armitage test and its generalizations is the somewhat arbitrary assignment of scores. However Snedecor and Cochran [20], page 246, report that moderate differences in the conclu- systems usually would not produce substantial differences in the conclusions from the analyses. We have used two sets of scores that seem natural-namely group indexes and logarithmic concentration. Since the concentrations were selected to be approximately equally spaced in the log domain, one would expect approximately the same results with both sets of scores.

We present below applications of the Cochran-Armitage test to the mortality responses in LeBlanc's Tests A and B and in Goulden's Isophorone Test. In LeBlanc's Test A we carried out separate comparisons, using the water control and using the solvent control groups as standards.

G1. LeBlanc Test A - Comparison With Water Control Group - Scores are Group Indices

We adjust for the effects of beaker to beaker heterogeneity within groups by utilizing the sample sizes and numbers of responses shown in Table V.3. The estimate of $\bar{p}$ is 88.1/301.8 = 0.292. The n_i's are 59.2 except for group 6 and $n_6 = 5.8$. The scores assigned to the groups are $X_1 = 1$, $X_3 = 2$, $X_4 = 3$, $X_5 = 4$, $X_6 = 5$, and $X_7 = 6$. Group 2, the solvent control group, is excluded from the analysis.

We fitted a cubic polynomial with weighted least squares using weights $w_i = n_i/\bar{p}\bar{q}$, as described above. The output showing the details of the straight line submodel fit and goodness of fit tests for the linear, quadratic, cubic, and departures from cubic models is displayed in Figure IX.14. The estimate, $\hat{\beta}_1$, of the slope β, is 0.15720 and appears at the upper right of the figure. A standard error estimate, 0.06390, appears to the right of $\hat{\beta}$, but this value must be modified before being used. Namely the model fitted by the weighted least squares program is

$$p_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2 w_i)$$

Under binomial theory, $\sigma^2 = 1$. However the algorithm estimates σ^2 by the residual mean square error, namely $\hat{\sigma}^2 = 17.66954$ and incorporates $\hat{\sigma}$ in all the standard error estimates and thus in all the t-ratios. We normalize the standard error estimate back to $\sigma=1$ by dividing it by $\hat{\sigma}$. Thus, under binomial theory

$$\text{std err } (\hat{\beta}_1) = 0.06390/(17.66954)^{1/2} = 0.01520 \quad .$$

The appropriate "t-value" is

$$t = 0.15720/0.01520 = 10.341 \quad .$$

This value can also be calculated by multiplying the stated t-value by $\hat{\sigma}$. Namely $(2.46)(17.66954)^{1/2} = 10.341$. To test H_0: $\beta_1 = 0$ versus H_1: $\beta_1 > 0$ we compare t with the percentiles of a standard normal distribution. A one tailed test is appropriate. Thus $\hat{\beta}_1$ is significant at the $\alpha = 0.0000$ level and so we strongly reject H_0.

The sums of squares at the bottom of Figure IX.14 can also be used to test hypotheses about the homogeneity or trend in the p_i's. The sum of squares corresponding to degree 0 represents the variability explained by fitting a cubic polynomial to the $\hat{p}_i$'s over and above a constant term. (Note that the entires in the column labeled "Sum of Squares" are to be used for inferences rather than the entries in the column labeled "F" since under binomial theory the residual mean square is known to be 1). The sums of squares corresponding to degrees 1 and 2 represent the variability explained by fitting a cubic polynomial to the $\hat{p}_i$'s over and above linear and quadratic polynomials, respectively. The residual sum of squares represents departures from a cubic polynomial. These sums of squares can be used to form various test statistics:

Degree 0 + Resid = 176.964 + 0.670 = 177.634 with 3 + 2 = 5 d.f.

This represents the deviation from the model of homogeneous probabilities. That is, it is the usual Pearson chi square statistic. We see that this agrees with the chi square value obtained in subsection D1 and is highly significant.

Degree 0 - Degree 1 = 176.96388 - 70.00806 = 106.956 with 3 - 2 = 1 d.f.

This represents the variation explained by a linear trend term in the p_i's. This sum of squares is compared to the percentiles of a chi square distribution with 1 d.f. and is highly significant. This test for linear trend is essentially equivalent to the Cochran-Armitage test. Note that $(106.956)^{1/2} = 10.342$, which agrees with the t value calculated previously, except that it is a two tailed test rather than a one tailed test.

Degree 1 + Resid = 70.00806 + 0.67011 = 70.678 with 2 + 2 = 4 d.f.

This represents deviation from the model of straight line trends. It is highly significant. There is thus evidence of departure from a straight line trend. The above decomposition is that which is usually referred to in texts as $\chi^2_{tot} = \chi^2_{slope} + \chi^2_{linearity}$.

Degree 2 + Resid = 5.96232 + 0.67011 = 6.632 with 1 + 2 = 3 d.f.

This represents deviations from a quadratic model. It is significant at $\alpha = 0.08$ and so represents just marginal suggestion of departure from a quadratic trend. There is no evidence of departure from a cubic trend.

Since the Cochran-Armitage test is highly significant using all the groups, we delete group 7 from the data and recalculate the test statistics. The output appears in Figure IX.15. The meanings of the various estimates are directly analogous to those in the previous figure and so need not be explained again in detail. The weights need to be recalculated because $\bar{p}$ is different than before.

The slope estimate, $\hat{\beta}_1$, is now negative and very close to 0; namely $\hat{\beta}_1 = -0.00353$. The standard error is $0.03189/(3.14541)^{1/2} = 0.0180$. Thus the t-value is

$$t = -0.00353/0.0180 = -0.196$$

which is nonsignificant.

The decomposition of the Pearson chi square statistic is

$$\chi^2_{tot} = 8.733 + 0.742 = 9.475 \qquad \text{with 4 d.f.}$$

$$\chi^2_{slope} = 8.733 - 8.695 = 0.038 \qquad \text{with 1 d.f.}$$

$$\chi^2_{linearity} = 8.695 + 0.742 = 9.437 \qquad \text{with 3 d.f.}$$

The total chi square is marginally significant ($\alpha = 0.05$). The slope chi square is not significant. The departure from linearity chi square is significant ($\alpha = 0.024$). The departure from quadratic chi square is also marginal ($\alpha = 0.07$). We thus conclude that there is some statistical evidence of departure from homogeneity, but the heterogeneity is not linear. We see from Figure II.1 that the mortality rate is about constant in groups 1-5 and then increases in group 6. This is not a straight line trend, especially since the group 6 responses are discounted so heavily relative to those in the other groups.

Since the overall and the departure from linearity chi squares are significant, we delete group 6 from the data and recalculate the test statistics. The output appears in Figure IX.16.

The slope estimate is again negative and very close to 0; namely $\hat{\beta}_1 = -0.02247$. The standard error is $0.00530/(0.08299)^{1/2} = 0.0184$. Thus the t-value is

$$t = -0.02247/0.0184 = -1.221$$

which is significant. (It is almost significant in the wrong direction.) The total chi square is nonsignificant ($\alpha = 0.35$), the slope chi square is

not significant ($\alpha = 0.22$), and the departure from linearity chi square is not significant ($\alpha = 0.92$). The process is thus stopped at this point and group 5 is declared to be the MATC group. This agrees with the results from the chi square test and from Williams' test.

The tests that were carried out in the previous analyses were based on the normal and the chi square distributions. A slight refinement of this procedure would be to base these inferences on the t and F distributions, using denominator degrees of freedom calculated from Table V.3. These degrees of freedom are 153, 123, and 120 for the three sets of comparisons. They are sufficiently large that use of the normal and chi square distributions makes little difference.

The Cochran-Armitage test was recalculated using the solvent control group for comparison instead of the water control group. The results were similar to those above and are not shown. Group 5 was declared to be the MATC group. The only difference in outcomes was that the Cochran-Armitage test was significant ($\alpha = 0.04$) after the data from group 7 were deleted.

G2. LeBlanc Test B - Comparison With Combined Water and Solvent Control Groups - Scores are Group Indices

We adjust for the effects of beaker to beaker heterogeneity within groups by using the sample sizes and numbers of responses shown in Table V.4. Since there was no statistically significant difference in the mortality rates in the solvent and water control groups, the responses in these two groups were combined for comparison with the treatment groups. The estimate of $\bar{p}$ for all the groups is 42.5/354.54 = 0.120. The n_i's are 57.6 except for group 7 and $n_7 = 8.96$. The scores assigned to the groups are $X_1=1$, $X_2=1$, $X_3=2$, $X_4=3$, $X_5=4$, $X_6=5$ and $X_7=6$.

The models fitted are the same as those discussed for Test A, namely cubic polynomials in the scores. The interpretations of the computer printouts are also the same as those for Test A and so need not be explained in detail.

The output showing the details of the straight line submodel fit and goodness of fit tests for the linear, quadratic, cubic, and departures from cubic models is displayed in Figure IX.17. The slope estimate is $\hat{\beta}_1 = 0.01744$. The estimated standard error is $0.02761/(6.24926)^{1/2} = 0.0110$. Thus the t-value is

$$t = 0.01744/0.0110 = 1.579$$

Comparing t with the precentiles of the standard normal distribution, we see that t is marginally significant at (the one tailed) $\alpha = 0.057$.

The decomposition of the Pearson chi square statistic is

$$\chi^2_{tot} = 15.394 + 12.097 = 27.491 \qquad \text{with } 3 + 2 = 5 \text{ d.f.}$$

$$\chi^2_{slope} = 15.394 - 12.900 = 2.494 \qquad \text{with } 1 \text{ d.f.}$$

$$\chi^2_{linearity} = 12.9000 + 12.097 = 24.997 \qquad \text{with } 2 + 2 = 4 \text{ d.f.}$$

The total chi square and departure from linearity chi square are not highly significant. The slope chi square is at best borderline ($\alpha = 0.11$), reflecting the marginally significant slope discussed above. Thus there is marginal statistical evidence of departures from homogeneity of mortality rates across groups. The sums of squares in Figure IX.17 suggest that the trends in the p_i's are reflected mostly in the quadratic component and in departures from a cubic polynomial (e.g. fourth degree term?). We see from Figure II.3 that this behavior is due to the substantial increase in mortality rate in group 7 after being relatively constant in groups 1-6. This resembles either quadratic or quartic trend.

Since the Cochran-Armitage statistic is marginally significant, the departure from linearity chi square is highly significant, and Figure II.3 reveals a sharp upward trend in mortality rates at group 7 we continue the process. We delete group 7 from the data and recalculate the test statistics. The output appears in Figure IX.18. Note that the weights for this analysis differ from those used before, since the estimate of $\bar{p}$, based on groups 1-6 is 36.9/345.6 = 0.107. The models fitted and the interpretations of the computer printouts are the same as before. The slope estimate is $\hat{\beta}_1 = 0.0000$. The estimated standard error is $0.01552/(1.939)^{1/2} = 0.011$. The t-value is of course 0. Thus there is no linear trend whatsoever in the mortality rates in groups 1-6.

The decomposition of the Pearson chi square statistic is

$$\chi^2_{tot} = 5.28888 + 0.52815 = 5.817 \qquad \text{with } 4 \text{ d.f.}$$

$$\chi^2_{slope} = 5.28888 - 5.28888 = 0 \qquad \text{with } 1 \text{ d.f.}$$

$$\chi^2_{linearity} = 5.28888 + 0.52815 = 5.817 \qquad \text{with } 3 \text{ d.f.}$$

The total chi square and slope chi square statistics are each nonsignificant ($\alpha = 0.21$ and $\alpha = 1.0$ respectively). The departure from linearity chi square is perhaps marginal ($\alpha = 0.12$). Breaking this chi square up into quadratic, cubic, and residual terms we see that the cubic component dominates the other two and is significant at the $\alpha = 0.04$ level. This a marginal level of significance and might simply be the result of selection. Since there is a nonsignificant total chi square, no linear component to the trend, a nonsignificant quadratic component ($\alpha = 0.28$), a marginal cubic component, and no monotone trend evident in Figure II.3, we conclude that there is no discernible upward trend in mortality rates in groups 1-6 and we stop the process. Group 6 is declared to be the MATC group. This agrees with the conclusions from the chi square test and from Williams' test.

G3. LeBlanc Test B - Comparison With Combined Water and Solvent Control Groups - Scores are Log (Concentration)

We repeat the analyses carried out in paragraph G2 above, but basing group scores on log (concentration) rather than on group indices. Otherwise everything else stays the same. The aim of this reanalysis is to determine if the use of these two different sets of scores leads to different conclusions.

The output showing the details of the straight line submodel fit and goodness of fit tests for the linear, quadratic, cubic, and departures from cubic models is displayed in Figure IX.19. The slope estimate is $\hat{\beta}_1 = 0.00743$. The estimated standard error is $0.04346/(6.82295)^{1/2} = 0.0166$. Thus the t-value is

$$t = 0.00743/0.0166 = 0.447$$

This value of t is nonsignificant ($\alpha \approx 0.33$) based on the upper tail of the normal distribution.

The decomposition of the Pearson chi square statistic is

$\chi^2_{tot} = 20.171 + 7.320 = 27.491$ with 3 + 2 = 5 d.f.

$\chi^2_{slope} = 20.171 - 19.972 = 0.199$ with 1 d.f.

$\chi^2_{linearity} = 19.972 + 7.320 = 27.292$ with 4 d.f.

The total chi square and departure from linearity chi square are each highly significant. The slope chi square is nonsignificant ($\alpha = 0.66$). Comparing Figures II.3 and II.4 we see that the linear trend is steeper when scores are based on indices rather than log (concentration) since the control group is then much less separated from the treatment groups. Based on the sums of squares at the bottom of Figure IX.19 we see that the major components of the departure from linearity chi square are the quadratic ($\chi^2 = 14.82$) and departure from cubic ($\chi^2 = 7.32$). We see from Figure II.4 that the trend in mortality rates with increasing concentration resembles a quadratic or a quartic curve.

Since the departure from linearity chi square is highly significant and Figure II.4 reveals a sharply upward trend in mortality rates at group 7 we continue the process. We delete group 7 from the data and recalculate the test statistics. The output appears in Figure IX.20. The slope estimate is now negative, $\hat{\beta}_1 = -0.01111$. The estimated standard error is $\hat{\sigma} = 0.02170/(1.78317)^{1/2} = 0.016$. The t-value is

$$t = -0.01111/0.016 = -0.684$$

This is nonsignificant ($\alpha = 0.75$ is the one tailed level). Thus there is no statistical evidence of linear trend in the mortality rates in groups 1-6.

The decomposition of the Pearson chi square statistic is

$$\chi^2_{tot} = 5.817 \qquad \text{with 4 d.f.}$$

$$\chi^2_{slope} = 0.467 \qquad \text{with 1 d.f.}$$

$$\chi^2_{linearity} = 5.350 \qquad \text{with 3 d.f.}$$

The total, slope, and departure from linearity chi squares are each non-significant ($\alpha = 0.21$, $\alpha = 0.49$, and $\alpha = 0.15$ respectively). The largest component of chi square is the quadratic component ($\chi^2 = 2.524$) which is at best marginal ($\alpha = 0.11$). Furthermore no monotone trend is evident in Figure II.4. We conclude that there is no discernible upward trend in mortality rates in groups 1-6 and we stop the process. Group 6 is declared to be the MATC group. This agrees with the conclusions from the chi square test, from Williams' test, and from the Cochran-Armitage test using group indices as scores.

G4. Goulden Isophorone Test - Scores are Group Indices

We consider a second data set to compare the results of applying the Cochran-Armitage test with scores based on group indices to the results with scores based on log concentration. We confine attention to the three beakers per group containing multiple daphnids (5 daphnids per beaker). Since there is no evidence of beaker to beaker heterogeneity within groups, no adjustments of sample sizes were carried out. There is just one control group.

The details of the straight line submodel fit and goodness of fit tests are displayed in Figure IX.21. The interpretation of the entries in the output are the same as those discussed previously for the LeBlanc data. The slope estimate is $\hat{\beta}_1 = 0.1486$ and its estimated standard error is $0.04395/(2.36554)^{1/2} = 0.0286$. Thus the t-value is

$$t = 0.1486/0.0286 = 5.200$$

This value is very highly significant, based on the upper tail of the normal distribution. The goodness of fit chi squares reflect the highly statistically significant linear trend component, as well as significant ($\alpha = 0.01$) departures from linearity. The trend in mortality rates is clearly evident in Figure II.8.

Since the Cochran-Armitage statistic and the depature from linearity chi square are both highly significant and Figure II.8 reveals a trend in mortality rates, we continue the process. We delete group 6 from the data and recalculate the test statistics. The output appears in Figure IX.22. The slope has been reduced from 0.1486 to 0.0933 but this is still highly significant. The t-value is 2.857, which is significant at the $\alpha = 0.002$ level. The departure from linearity chi square is marginal ($\alpha = 0.11$).

Since the Cochran-Armitage statistic is significant and Figure II.8 reveals a trend even after deleting group 6, we continue the process. We delete group 5 from the data and recalculate the test statistics. The slope estimate has now been reduced to 0.02 and is no longer significant ($\alpha = 0.29$). The departure from linearity chi square is also nonsignificant. ($\alpha = 0.49$). We thus stop the process and declare group 4 to be the MATC group. This agrees with the results of the chi square test, the measure of association test, and Williams' test.

G5. Goulden Isophorone Test - Scores are Log (Concentration)

We repeat the analyses carried out in paragraph G4 above using the same data but with scores based on log (concentration). In particular since the control group is at (nominal) concentration 0, we define the scores to be log (1 + concentration).

The computer outputs corresponding to all the groups, group 6 omitted, and groups 5 and 6 omitted are displayed in Figure IX.24, 25, and 26 respectively. Based on all the data, both the Cochran-Armitage test and the departure from linearity chi square test are significant ($\alpha = 0.01$, $\alpha = 0.0000$ respectively). After deleting group 6, both these statistics are still significant ($\alpha = 0.025$, $\alpha = 0.0002$ respectively). After deleting groups 5 and 6, neither of these statistics are significant ($\alpha = 0.41$, $\alpha = 0.18$ respectively). We thus stop the process at this point and declare group 4 to be the MATC group. This result agrees with that obtained based on using the group indices as scores and with those obtained based on the chi square, measure of association, and Williams' tests.

In summary we have illustrated the Cochran-Armitage test and generalizations on three data sets. On two of the data sets we used scores based both on group indices and on log (concentration). While there were some relatively minor differences in detailed results, both sets of scores led to the same conclusion about the MATC group in both data sets. This suggests that the results of the Cochran-Armitage procedure are not too sensitive to moderate differences in scores, such as in the two sets we used. Thus the Cochran-Armitage test appears to be a reasonable procedure to use for the comparison of mortality rates in chronic daphnia tests. This matter needs further empirical or theoretical study utilizing a number of other data sets and choices of scores.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ****************************************

21-DAY ADJUSTED MORTALITY, COMPARE TO WATER CONTROL, OMIT SOLVENT CONTROL

                    ****************************************

GROUP  1                CONTINGENCY TABLE

                        LIVE          DEAD          TOTAL
          TRT    1      50.300         8.900        59.200
          TRT    3      51.800         7.400        59.200
          TRT    4      54.000         5.200        59.200
          TRT    5      54.000         5.200        59.200
          TRT    6       2.900         2.900         5.800
          TRT    7        .700        58.500        59.200

          TOTAL LIVE  =   213.700  , TOTAL DEAD  =    88.100  , GRAND TOTAL =   301.800

                    EXPECTED FREQUENCIES                        CUTOFF =    .00100

                        LIVE          DEAD          TOTAL
          TRT    1      41.919        17.281        59.200
          TRT    3      41.919        17.281        59.200
          TRT    4      41.919        17.281        59.200
          TRT    5      41.919        17.281        59.200
          TRT    6       4.107         1.693         5.800
          TRT    7      41.919        17.281        59.200

          TOTAL LIVE  =   213.700  , TOTAL DEAD  =    88.100  , GRAND TOTAL =   301.800

  1   INDEPENDENT TESTS COMBINED
    ASYMPTOTIC CHI SQUARED USED
      OBSERVED CHI SQUARED =        177.63399
      OBSERVED SIGNIFICANCE LEVEL = AI =        0.0000000000
```

Figure IX.1. Chi square test of homogeneity across groups - LeBlanc Test A - 21 day mortality-water control group - EXAX output.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ****************************************
21-DAY ADJUSTED MORTALITY, USE WATER CONTROL, OMIT SOLVENT CONTROL, DROP GRP 7
                    ****************************************

GROUP  1              CONTINGENCY TABLE

                      LIVE          DEAD          TOTAL
        TRT    1      50.300         8.900        59.200
        TRT    3      51.800         7.400        59.200
        TRT    4      54.000         5.200        59.200
        TRT    5      54.000         5.200        59.200
        TRT    6       2.900         2.900         5.800

        TOTAL LIVE   =   213.000  , TOTAL DEAD   =    29.600  , GRAND TOTAL =   242.600

                  EXPECTED FREQUENCIES                        CUTOFF =    .00100

                      LIVE          DEAD          TOTAL
        TRT    1      51.977         7.223        59.200
        TRT    3      51.977         7.223        59.200
        TRT    4      51.977         7.223        59.200
        TRT    5      51.977         7.223        59.200
        TRT    6       5.092          .708         5.800

        TOTAL LIVE   =   213.000  , TOTAL DEAD   =    29.600  , GRAND TOTAL =   242.600

  1  INDEPENDENT TESTS COMBINED
     ASYMPTOTIC CHI SQUARED USED
        OBSERVED CHI SQUARED =          9.47474
        OBSERVED SIGNIFICANCE LEVEL = AI =          .0502689344
```

Figure IX.2. Chi square test of homogeneity across groups - LeBlanc Test A - 21 day mortality - water control group - group 7 omitted - EXAX2 output.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ****************************************
21-DAY ADJUSTED MORTALITY, COMPARE TO SOLVENT CONTROL, OMIT WATER CONTROL
                    ****************************************

GROUP  1                CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
        TRT   2     57.000         2.200        59.200
        TRT   3     51.800         7.400        59.200
        TRT   4     54.000         5.200        59.200
        TRT   5     54.000         5.200        59.200
        TRT   6      2.900         2.900         5.800
        TRT   7       .700        58.500        59.200

        TOTAL LIVE  =   220.400  , TOTAL DEAD  =    81.400  , GRAND TOTAL =   301.800

                EXPECTED FREQUENCIES                        CUTOFF =   .00100

                    LIVE          DEAD          TOTAL
        TRT   2     43.233        15.967        59.200
        TRT   3     43.233        15.967        59.200
        TRT   4     43.233        15.967        59.200
        TRT   5     43.233        15.967        59.200
        TRT   6      4.236         1.564         5.800
        TRT   7     43.233        15.967        59.200

        TOTAL LIVE  =   220.400  , TOTAL DEAD  =    81.400  , GRAND TOTAL =   301.800

1   INDEPENDENT TESTS COMBINED
  ASYMPTOTIC CHI SQUARED USED
     OBSERVED CHI SQUARED =        199.13679
     OBSERVED SIGNIFICANCE LEVEL = AT =        0.0000000000
```

Figure IX.3. Chi square test of homogeneity across groups - LeBlanc Test A - 21 day mortality - solvent control group - EXAX2 output.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ****************************************

21-DAY ADJUSTED MORTALITY, USE SOLVENT CONTROL, OMIT WATER CONTROL, DROP GRP 7

                    ****************************************


GROUP  1                CONTINGENCY TABLE

                      LIVE          DEAD          TOTAL
        TRT    2      57.000         2.200        59.200
        TRT    3      51.800         7.400        59.200
        TRT    4      54.000         5.200        59.200
        TRT    5      54.000         5.200        59.200
        TRT    6       2.900         2.900         5.800

        TOTAL LIVE     =    219.700  , TOTAL DEAD     =     22.900  , GRAND TOTAL =    242.600

                EXPECTED FREQUENCIES                              CUTOFF =     .00100

                      LIVE          DEAD          TOTAL
        TRT    2      53.612         5.588        59.200
        TRT    3      53.612         5.588        59.200
        TRT    4      53.612         5.588        59.200
        TRT    5      53.612         5.588        59.200
        TRT    6       5.253          .547         5.800

        TOTAL LIVE     =    219.700  , TOTAL DEAD     =     22.900  , GRAND TOTAL =    242.600

  1   INDEPENDENT TESTS COMBINED .
    ASYMPTOTIC CHI SQUARED USED
      OBSERVED CHI SQUARED =        14.13889
      OBSERVED SIGNIFICANCE LEVEL = AI =        .0068647252
```

Figure IX.4. Chi square test of homogeneity across groups - LeBlanc Test A - 21 day mortality - solvent control group - group 7 omitted - EXAX2 output.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST A

21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA

1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)

4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))

GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),

MEASURED CONC (0,7,14,21)

                    ****************************************

21-DAY ADJUSTED MORTALITY, USE SOLVENT CONTROL, OMIT WATER CONTROL, DROP GRP 6&7

                    ****************************************


GROUP  1              CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
        TRT   2      57.000         2.200        59.200
        TRT   3      51.800         7.400        59.200
        TRT   4      54.000         5.200        59.200
        TRT   5      54.000         5.200        59.200

        TOTAL LIVE   =   216.800  , TOTAL DEAD   =    20.000  , GRAND TOTAL =   236.800

                  EXPECTED FREQUENCIES                        CUTOFF =    .00100

                    LIVE          DEAD          TOTAL
        TRT   2      54.200         5.000        59.200
        TRT   3      54.200         5.000        59.200
        TRT   4      54.200         5.000        59.200
        TRT   5      54.200         5.000        59.200

        TOTAL LIVE   =   216.800  , TOTAL DEAD   =    20.000  , GRAND TOTAL =   236.800

  1   INDEPENDENT TESTS COMBINED
    ASYMPTOTIC CHI SQUARED USED
      OBSERVED CHI SQUARED =          2.98840
      OBSERVED SIGNIFICANCE LEVEL = AI =        .3934173576
```


Figure IX.5. Chi square test of homogeneity across groups - LeBlanc test A - 21 day mortality - solvent control group - groups 6 and 7 omitted - EXAX2 output.

```
INVESTIGATOR: EG AND G BIONOMICS -- TEST B
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. PERT. (7,14,21),
MEASURED CONC (0,7,14,21)
```

****

21-DAY ADJUSTED MORTALITY, WATER+SOLVENT CONTROLS POOLED

****

GROUP 1 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	192	100.000	15.200	115.200
TRT	3	54.700	2.900	57.600
TRT	4	54.000	3.600	57.600
TRT	5	48.200	9.400	57.600
TRT	6	51.800	5.800	57.600
TRT	7	3.400	5.560	8.960

TOTAL LIVE = 312.100 , TOTAL DEAD = 42.460 , GRAND TOTAL = 354.560

EXPECTED FREQUENCIES CUTOFF = .00100

		LIVE	DEAD	TOTAL
TRT	192	101.404	13.796	115.200
TRT	3	50.702	6.898	57.600
TRT	4	50.702	6.898	57.600
TRT	5	50.702	6.898	57.600
TRT	6	50.702	6.898	57.600
TRT	7	7.887	1.073	8.960

TOTAL LIVE = 312.100 , TOTAL DEAD = 42.460 , GRAND TOTAL = 354.560

```
1  INDEPENDENT TESTS COMBINED
ASYMPTOTIC CHI SQUARED USED
   OBSERVED CHI SQUARED =          27.13174
   OBSERVED SIGNIFICANCE LEVEL = AI =          .0000537696
```

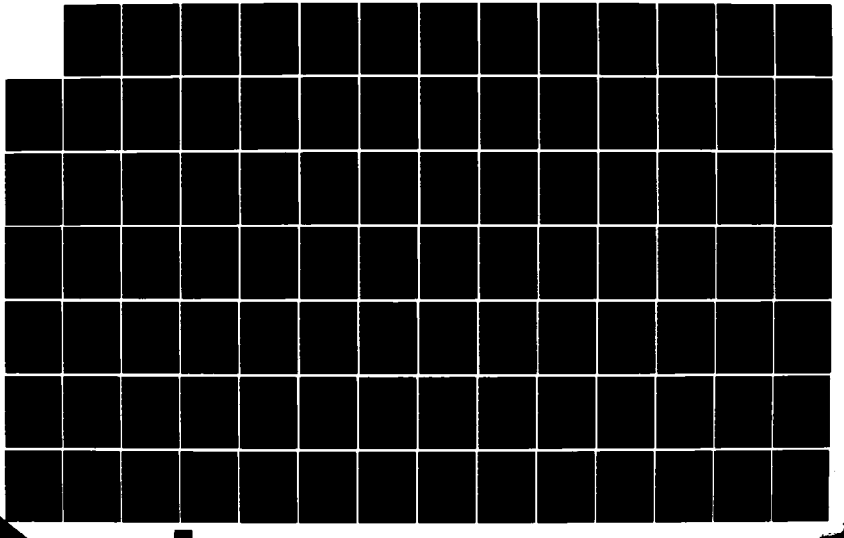
Figure IX.6. Chi square test of homogeneity across groups - LeBlanc test B - 21 day mortality - solvent and water control groups combined - EXAX2 output

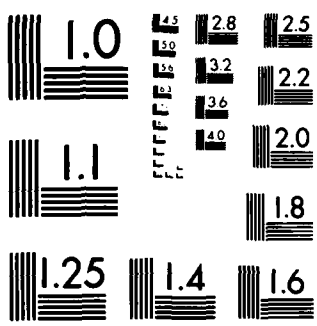
```
INVESTIGATOR: EG AND G BIONOMICS -- TEST B
21 DAY FLOW THROUGH TEST WITH DAPHNIA MAGNA
1 CONTROL GROUP (1), 1 SOLVENT GROUP (2), 5 TRX GROUPS (3-7)
4 BEAKERS PER GROUP--FORMAT (I1,1X,A1,2F7.4,4I4,3F5.1,4(A1,F6.4))
GRP, REPL, CONC (MG/L) NO. TESTED, NO. LIVE (7,14,21), CUM. FERT. (7,14,21),
MEASURED CONC (0,7,14,21)
                    ****************************************
21-DAY ADJUSTED MORTALITY, WATER+SOLVENT CONTROLS POOLED, DROP OFF GRP 7
                    ****************************************

GROUP  1              CONTINGENCY TABLE

                      LIVE          DEAD          TOTAL
      TRT   #2      100.000        15.200        115.200
      TRT   3        54.700         2.900         57.600
      TRT   4        54.000         3.600         57.600
      TRT   5        48.200         9.400         57.600
      TRT   6        51.800         5.800         57.600

      TOTAL LIVE    =    308.700  , TOTAL DEAD     =     36.900   , GRAND TOTAL =    345.600

                EXPECTED FREQUENCIES                          CUTOFF =    .00100

                      LIVE          DEAD          TOTAL
      TRT   #2      102.900        12.300        115.200
      TRT   3        51.450         6.150         57.600
      TRT   4        51.450         6.150         57.600
      TRT   5        51.450         6.150         57.600
      TRT   6        51.450         6.150         57.600

      TOTAL LIVE    =    308.700  , TOTAL DEAD     =     36.900   , GRAND TOTAL =    345.600

1   INDEPENDENT TESTS COMBINED
  ASYMPTOTIC CHI SQUARED USED
     OBSERVED CHI SQUARED =          5.81702
     OBSERVED SIGNIFICANCE LEVEL = AI =        .2132361505
```

Figure IX.7. Chi square test of homogeneity across groups - LeBlanc test B - 21 day mortality - solvent and water control groups combined - group 7 omitted -EXAX2 output

AD-A152 070 DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA(U) BATTELLE COLUMBUS LABS OH P I FEDER DEC 81 BATT-G-7673 DAMD17-80-C-0165 4/5

UNCLASSIFIED F/G 6/20 NL





GOULDEN SURVIVAL DATA, REPS 8-10 ONLY (5 INDIV/BEAKER)

****

DAPHNIA 21-DAY MORTALITY

****

GROUP 1 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	13.000	2.000	15.000
TRT	2	14.000	1.000	15.000
TRT	3	14.000	1.000	15.000
TRT	4	12.000	3.000	15.000
TRT	5	7.000	8.000	15.000
TRT	6	2.000	13.000	15.000

TOTAL LIVE = 62.000 , TOTAL DEAD = 28.000 , GRAND TOTAL = 90.000

EXPECTED FREQUENCIES CUTOFF = .00100

		LIVE	DEAD	TOTAL
TRT	1	10.333	4.667	15.000
TRT	2	10.333	4.667	15.000
TRT	3	10.333	4.667	15.000
TRT	4	10.333	4.667	15.000
TRT	5	10.333	4.667	15.000
TRT	6	10.333	4.667	15.000

TOTAL LIVE = 62.000 , TOTAL DEAD = 28.000 , GRAND TOTAL = 90.000

1 INDEPENDENT TESTS COMBINED
ASYMPTOTIC CHI SQUARED USED
OBSERVED CHI SQUARED = 36.49770
OBSERVED SIGNIFICANCE LEVEL = AI = .0000007551

Figure IX.8. Chi square test of homogeneity across groups - Goulden isophorone test - 21 day mortality - multiply housed daphnids - EXAX2 output.

GOULDEN SURVIVAL DATA, REPS 8-10 ONLY, DROP GROUP 6

****

DAPHNIA 21-DAY MORTALITY

****

GROUP 1 CONTINGENCY TABLE

		LIVE	DEAD	TOTAL
TRT	1	13.000	2.000	15.000
TRT	2	14.000	1.000	15.000
TRT	3	14.000	1.000	15.000
TRT	4	12.000	3.000	15.000
TRT	5	7.000	8.000	15.000

TOTAL LIVE = 60.000 , TOTAL DEAD = 15.000 , GRAND TOTAL = 75.000

EXPECTED FREQUENCIES CUTOFF = .00100

		LIVE	DEAD	TOTAL
TRT	1	12.000	3.000	15.000
TRT	2	12.000	3.000	15.000
TRT	3	12.000	3.000	15.000
TRT	4	12.000	3.000	15.000
TRT	5	12.000	3.000	15.000

TOTAL LIVE = 60.000 , TOTAL DEAD = 15.000 , GRAND TOTAL = 75.000

1 INDEPENDENT TESTS COMBINED
ASYMPTOTIC CHI SQUARED USED
OBSERVED CHI SQUARED = 14.16667
OBSERVED SIGNIFICANCE LEVEL = AI = .0067816896

Figure IX.9. Chi square test of homogeneity across groups - Goulden isophorone test - 21 day mortality - multiply housed daphnids - EXAX2 output - group 6 omitted.

```
GOULDEN SURVIVAL DATA, REPS 8-10 ONLY, DROP GROUPS 5-6

*****************************************

DAPHNIA 21-DAY MORTALITY

*****************************************

GROUP  1                CONTINGENCY TABLE

                    LIVE          DEAD          TOTAL
      TRT    1      13.000         2.000        15.000
      TRT    2      14.000         1.000        15.000
      TRT    3      14.000         1.000        15.000
      TRT    4      12.000         3.000        15.000

      TOTAL LIVE    =     53.000  , TOTAL DEAD    =      7.000  , GRAND TOTAL =     60.000

                EXPECTED FREQUENCIES                          CUTOFF =    .00100

                    LIVE          DEAD          TOTAL
      TRT    1      13.250         1.750        15.000
      TRT    2      13.250         1.750        15.000
      TRT    3      13.250         1.750        15.000
      TRT    4      13.250         1.750        15.000

      TOTAL LIVE    =     53.000  , TOTAL DEAD    =      7.000  , GRAND TOTAL =     60.000

  1   INDEPENDENT TESTS COMBINED
    ASYMPTOTIC CHI SQUARED USED
      OBSERVED CHI SQUARED =          1.77898
      OBSERVED SIGNIFICANCE LEVEL = AI =          .6195207011
```

Figure IX.10. Chi square test of homogeneity across groups - Goulden isophorone test - 21 day mortality - multiply housed daphnids - EXAX2 output - groups 5 and 6 omitted.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 4)

TABLE NO. 1 GRP (VAR 1) VS SURVIVAL(VAR 2)

CELL FREQUENCY COUNTS

			SURVIVAL(VAR 2)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	GRP1	1.00	13	2	15
	GRP2	2.00	14	1	15
	GRP3	3.00	14	1	15
	GRP4	4.00	12	3	15
	GRP5	5.00	7	8	15
	GRP6	6.00	2	13	15
	TOTAL		62	28	90

MINIMUM NONZERO EXPECTED VALUE IS 4.667
----- ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	36.498	5	.0000				

STATISTIC	VALUE	ASE1	T-VALUE	DEP.	STATISTIC	VALUE	ASE1	T-VALUE	DEP.
GAMMA	.748	.102	5.730		STUART'S TAU-C	.578	.101	5.730	
					KENDALL'S TAU-B	.483	.076	5.730	
SOMER'S D	.347	.061	5.730	2	SOMER'S D	.674	.102	5.730	1

PERCENTAGES OF THE ROW TOTALS

			SURVIVAL(VAR 2)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	GRP1	1.00	86.67	13.33	100.00
	GRP2	2.00	93.33	6.67	100.00
	GRP3	3.00	93.33	6.67	100.00
	GRP4	4.00	80.00	20.00	100.00
	GRP5	5.00	46.67	53.33	100.00
	GRP6	6.00	13.33	86.67	100.00
	TOTAL		68.89	31.11	100.00

Figure IX.11. BMDP1F output from one sided measure of association test - Goulden isophorone test - 21 day mortality.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 4)

TABLE NO. 1 GRP (VAR 1) VS SURVIVAL(VAR 2)

CELL FREQUENCY COUNTS

			SURVIVAL(VAR 2)		
		EQ./EQ.	ALIVE 1.00	DEAD 2.00	TOTAL
GRP (VAR 1)	GRP1	1.00	13	2	15
	GRP2	2.00	14	1	15
	GRP3	3.00	14	1	15
	GRP4	4.00	12	3	15
	GRP5	5.00	7	8	15
	TOTAL		60	15	75

MINIMUM NONZERO EXPECTED VALUE IS 3.000
***** ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	14.167	4	.0068				

STATISTIC	VALUE	ASE1	T-VALUE	DEP.	STATISTIC	VALUE	ASE1	T-VALUE	DEP.
GAMMA	.557	.183	2.583		STUART'S TAU-C	.299	.116	2.583	
					KENDALL'S TAU-B	.295	.105	2.583	
SOMER'S D	.187	.072	2.583	2	SOMER'S D	.467	.162	2.583	1

Figure IX.12. BMDP1F output from one sided measure of association test - Goulden isophorone test - 21 day mortality - group 6 omitted.

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 4)

TABLE NO. 1 GRP (VAR 1) VS SURVIVAL(VAR 2)

CELL FREQUENCY COUNTS

			SURVIVAL(VAR 2)		
			ALIVE	DEAD	
		EQ./EQ.	1.00	2.00	TOTAL
GRP	GRP1	1.00	13	2	15
(VAR 1)	GRP2	2.00	14	1	15
	GRP3	3.00	14	1	15
	GRP4	4.00	12	3	15
	TOTAL		53	7	60

MINIMUM NONZERO EXPECTED VALUE IS 1.750
----- ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	1.779	3	.6195				

STATISTIC	VALUE	ASE1	T-VALUE	DEP.	STATISTIC	VALUE	ASE1	T-VALUE	DEP.
GAMMA	.160	.334	.472		STUART'S TAU-C	.050	.106	.472	
					KENDALL'S TAU-B	.064	.134	.472	
SOMER'S D	.033	.071	.472	2	SOMER'S D	.121	.254	.472	1

Figure IX.13. BMDP1F output from one sided measure of association test - Goulden isophorone test - 21 day mortality - groups 5 and 6 omitted.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	11.15438	4.20352	2.65	-.21657	.23413	-.93
1	10.34194	4.20352	2.46	.15720	.06390	2.46

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 17.66954 (D.F. = 4)
MULTIPLE R-SQUARE = .99916

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	176.96388	3	58.98796	176.05	.01
1	70.00806	2	35.00403	104.47	.01
2	5.96232	1	5.96232	17.80	.05
RESID.	.67011	2	.33506		

Figure IX.14. Cochran-Armitage test of homogeneity across groups - LeBlanc test A - 21 day mortality - water control group. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	5.80633	1.77353	3.27	.13104	.08973	1.46
1	-.19623	1.77353	-.11	-.00353	.03189	-.11

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2_A$ = 3.14541 (D.F. = 3)
MULTIPLE R-SQUARE = .99945

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	8.73319	3	2.91106	3.93	.35
1	8.69468	2	4.34734	5.86	.28
2	4.54207	1	4.54207	6.13	.24
RESID.	.74155	1	.74155		

Figure IX.15. Cochran-Armitage test of homogeneity across groups - LeBlanc test A - 21 day mortality - group 7 omitted - water control group. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	5.48572	.28808	19.04	.16892	.01450	11.65
1	-1.22205	.28808	-4.24	-.02247	.00530	-4.24

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = .08239 (D.F. = 2)
MULTIPLE R-SQUARE = .99999

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	1.58838	2	.79419	11.19	.21
1	.09498	1	.09498	1.34	.45
RESID.	.07100	1	.07100		

Figure IX.16. Cochran-Armitage test of homogeneity across groups - LeBlanc test A - 21 day mortality - groups 6 and 7 omitted - water control group. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.94897	2.49985	2.78	.07189	.08733	.82
1	1.57925	2.49985	.63	.01744	.02761	.63

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 6.24926 (D.F. = 4)
MULTIPLE R-SQUARE = .99931

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	15.39443	3	5.13148	.85	.58
1	12.90040	2	6.45020	1.07	.48
2	1.77737	1	1.77737	.29	.64
RESID.	12.09665	2	6.04832		

Figure IX.17. Cochran-Armitage test of homogeneity across groups - LeBlanc test B - 21 day mortality - combined control groups. Scores are group indices.

DESIGNS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.42735	1.39248	4.62	.10677	.04741	2.25
1	.00000	1.39248	.00	.00000	.01552	.00

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$=.193900+01 (D.F. = 3)
MULTIPLE R-SQUARE = .99986

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	5.28888	3	1.76296	3.34	.38
1	5.28888	2	2.64444	5.01	.30
2	4.14105	1	4.14105	7.84	.22
RESID.	.52815	1	.52815		

Figure IX.18. Cochran-Armitage test of homogeneity across groups - LeBlanc test B - 21 day mortality - group 7 omitted - combined control groups. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.94897	2.61208	2.66	.13533	.10109	1.34
1	.44640	2.61208	.17	.00743	.04346	.17

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$=6.82235 (D.F. = 4)
MULTIPLE R-SQUARE = .39925

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	20.17129	3	6.72376	1.84	.37
1	19.97202	2	9.98601	2.73	.27
2	5.14738	1	5.14738	1.41	.36
RESID.	7.31978	2	3.65989		

Figure IX.19. Cochran-Armitage test of homogeneity across groups - LeBlanc test B - 21 day mortality - combined control groups. Scores are log (concentration).

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.42735	1.33535	4.81	.08321	.05108	1.63
1	-.68375	1.33535	-.51	-.01111	.02170	-.51

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 1.78317 (D.F. = 3)
MULTIPLE R-SQUARE = .99987

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	4.01419	3	1.33806	.74	.67
1	3.54668	2	1.77334	.98	.58
2	1.02295	1	1.02295	.57	.59
RESID.	1.80284	1	1.80284		

Figure IX.20. Cochran-Armitage test of homogeneity across groups - LeBlanc test B - 21 day mortality - group 7 omitted - combined control groups. Scores are log (concentration).

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.37536	1.53803	4.15	-.20889	.17115	-1.22
1	5.19957	1.53803	3.38	.14857	.04395	3.38

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 2.36554 (D.F. = 4)
MULTIPLE R-SQUARE = .99823

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	FAIL PROBABILITY
0	36.30020	3	12.10007	122.53	.01
1	9.26465	2	4.63232	46.91	.02
2	.00691	1	.00691	.07	.82
RESID.	.19750	2	.09875		

Figure IX.21. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - 21 day mortality. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	4.33013	1.41421	3.06	-.08000	.15319	-.52
1	2.85774	1.41421	2.02	.09333	.04619	2.02

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 2.00000 (D.F. = 3)
MULTIPLE R-SQUARE = .99804

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THIS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	14.16667	3	4.72222	INFINITY	
1	6.00000	2	3.00000	INFINITY	
2	.16667	1	.16667	INFINITY	
RESID.	.00000	1	.00000		

Figure IX.22. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - group 6 omitted - 21 day mortality. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	2.81505	.86252	3.26	.06667	.08756	.76
1	.53954	.86252	.63	.02000	.03197	.63

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = .74394 (D.F. = 2)
MULTIPLE R-SQUARE = .99894

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THIS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	1.74663	2	.87332	27.00	.13
1	1.45553	1	1.45553	45.00	.09
RESID.	.03235	1	.03235		

Figure IX.23. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - groups 5 and 6 omitted - 21 day mortality. Scores are group indices.

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	6.37536	2.38658	2.67	-.03594	.25216	-.14
1	3.70332	2.38658	1.55	.22547	.14530	1.55

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 5.69579 (D.F. = 4)
MULTIPLE R-SQUARE = .99575

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	35.79960	3	11.93320	34.19	.03
1	22.08504	2	11.04252	31.64	.03
2	6.82016	1	6.82016	19.54	.05
RESID.	.69810	2	.34905		

Figure IX.24. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - 21 day mortality. Scores are log (1 + concentration).

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	4.33013	1.85464	2.33	.04185	.17233	.24
1	1.96152	1.85464	1.06	.11407	.10785	1.06

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = 3.43979 (D.F. = 3)
MULTIPLE R-SQUARE = .99664

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	13.64113	3	4.54704	8.65	.24
1	9.79357	2	4.89678	9.32	.23
2	3.31934	1	3.31934	6.32	.24
RESID.	.52554	1	.52554		

Figure IX.25. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - group 6 omitted - 21 day mortality data. Scores are log (1 + concentration).

RESULTS FOR POLYNOMIAL OF DEGREE 1

POLYNOMIAL COEFFICIENTS	ORTHOGONAL POLYNOMIAL			POLYNOMIAL IN X		
DEGREE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE	REGRESSION COEFFICIENT	STANDARD ERROR	T VALUE
0	2.81505	.92877	3.03	.10183	.07081	1.44
1	.23184	.92877	.25	.01248	.05001	.25

RESIDUAL MEAN SQUARE = $\hat{\sigma}^2$ = .86261 (D.F. = 2)
MULTIPLE R-SQUARE = .99877

GOODNESS OF FIT TEST

FOR THE POLYNOMIAL OF EACH DEGREE, A TEST IS MADE FOR ADDITIONAL INFORMATION IN THE ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE. THE NUMERATOR SUM OF SQUARES FOR EACH OF THESE TESTS IS THE SUM OF SQUARES ATTRIBUTED TO ALL ORTHOGONAL POLYNOMIALS OF HIGHER DEGREE AND THE DENOMINATOR SUM OF SQUARES IS THE RESIDUAL SUM OF SQUARES FROM THE FIT TO THE HIGHEST DEGREE POLYNOMIAL (FIT TO ALL ORTHOGONAL POLYNOMIALS). A SIGNIFICANT F-STATISTIC THUS INDICATES THAT A HIGHER DEGREE POLYNOMIAL SHOULD BE CONSIDERED.

DEGREE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	TAIL PROBABILITY
0	1.30726	2	.65363	1.39	.51
1	1.25351	1	1.25351	2.66	.35
RESID.	.47171	1	.47171		

Figure IX.26. Cochran-Armitage test of homogeneity across groups - Goulden isophorone test - groups 5 and 6 omitted - 21 day mortality data. Scores are log (1 + concentration).

X. DOSE RESPONSE CURVE ESTIMATION-PROBIT ANALYSIS OF MORTALITY RATES

A. INTRODUCTION. DOSE RESPONSE CURVE ESTIMATION VS HYPOTHESIS TESTING. DESCRIPTIONS OF MODELS

An alternative approach to hypothesis testing and multiple comparisons for the determination of acceptable concentrations with respect to mortality is through fitting dose response models to the mortality data and estimating the concentrations, C_L, which result in incremental mortality of at most L over and above the background level. The problem of determining a safe concentration has been transformed from a hypothesis testing problem (determine which treatment group mortality rates are statistically significantly different from the control rate) to an estimation problem (obtain point estimates of and confidence intervals on C_L).

The hypothesis testing and dose response curve estimation problems are conceptually different and have different implications. A number of these differences were discussed by Feder and Collins [1], Section XIV. Hypothesis testing procedures provide information as to whether treatment group responses are statistically significantly different than control group responses. They say nothing about biologically significant differences. Namely the differences between treatment group and control group mortality rates may be highly statistically significant but yet biologically trivial, if the sample sizes are great enough. Conversely, biologically important differences may not be supported as being statistically significant if the sample sizes are too small. By contrast, dose response curve estimation procedures estimate those concentration levels that produce biologically significant changes in response. Biological significance is quantified by stating the increments from control rates that are considered to be important--for example 10 percent, 25 percent, etc., increases in mortality. Point and confidence interval inferences about concentrations associated with such differences are then constructed.

An additional conceptual difference between the two types of procedures is reflected in the effects on inferences of changes in sample sizes. With the classical hypothesis testing formulation the larger and more precise the experiment the more powerful will be the hypothesis test. Thus lower concentration levels will yield responses that are statistically significantly different then the control group response. The MATC will be decreased. Conversely the smaller the experiment is and the more variable the responses are, the greater will be the MATC. The effects of increased sample sizes and precision on inferences based on dose response curves are exactly the opposite. The larger and more precise the experiment, the tighter will be the confidence bounds on C_L (the concentration associated with an incremental response of L). Thus the lower confidence bound on C_L will be increased.

We feel that the latter situation is as it should be. The more extensive and the more precise the supporting evidence, the more liberal should be the determinations of safe concentration. Thus inferences about acceptable concentrations based on dose response curves depend on sample size in the manner they should whereas inferences based on hypothesis tests depend on sample size in a manner opposite to what they should.

Feder and Collins [1] discuss these conceptual differences at greater length. They illustrate a number of different dose response models and computer programs for fitting such models. In particular they illustrate probit models to describe trends in mortality rates with increases in either concentration or log (concentration). See Sections XIV, XV, XVI of Feder and Collins for details. Some of the topics discussed in those sections include:

- Fitting probit models in concentration or log concentration using the special purpose probit analysis module, PROC PROBIT, in the SAS statistical computing system [10].

- Fitting probit and logit models in concentration or log concentration using the general purpose SAS nonlinear regression model, PROC NLIN.

- Fitting adjustments for background mortality using Abbott's correction and alternatives.

- Fitting nonparametric dose response models that yield conservative lower confidence bounds on safe concentrations without assuming specific parametric forms for the response curve.

In this section we discuss fitting three parameter probit models to the 21-day mortality responses, using either concentration or log concentration as the independent variable, as appropriate. Background mortality is accounted for by Abbott's correction. See Finney [4] for a detailed description of this model and associated inferences.

Feder and Collins [1] discuss the tradeoffs involved in not adjusting for background mortality. That is, a two parameter probit model, incorporating the assumption of zero background mortality, could be fitted to the data. If background mortality is in fact present, then the estimated mortality rates based on these two parameter models would be biased downwards toward zero, especially at the low mortality end of the curve. However the standard errors of the estimates there would be reduced. Therefore if the background mortality is in fact nonzero but is not too far different from zero, the resulting increased precision might more than offset the downward biases, thus resulting in more accurate estimates.

This suggests that a special study should be made to determine how large the estimates of background mortality would need to be before background adjustments would be called for. The greater the quantity and the more precise the data, the less background variation that could be

tolerated before background adjustments were needed. In the absence of the results of such a detailed study, we have adopted here the somewhat arbitrary rule that if the estimated background mortality rate parameter in a three parameter probit fit is significantly different from zero at the 5 percent level, then we will retain it in the model, thereby adjusting for background.

The background mortality rates in the various data sets we have looked at are generally fairly high, often in excess of 10 percent. In the combined vehicle and water control groups in LeBlanc's tests A and B, 15 of 160 and 21 of 160 daphnids died, respectively. In Adams' test, 1 of 15 control daphnids died. In Chapman's test, 2 of 20 control daphnids died. In Goulden's test, 2 of 15 of the multiple housed control daphnids died.

We fitted three parameter probit models to the 21-day mortality data from LeBlanc's tests A and B and from Goulden's isophorone test. In each case the background rate was significant at the 5 percent level; we thus retained these corrections for background.

All probit fits were carried out using the general purpose BMDP nonlinear regression program, BMDPAR. See Dixon and Brown [7] for a detailed description of the computer program. See Jennrich and Moore [25] for a discussion of the theory underlying the use of nonlinear regression methods to perform maximum likelihood probit analysis fits. Although using nonlinear regression programs to fit probit dose response models is a bit more fussy than using special purpose probit analysis programs, there are a number of advantages to this approach. These include:

- Program availability. A user may not have a special purpose computer program accessible that will fit three parameter probit models. In particular, SAS PROC PROBIT is available only to SAS users. By contrast, any nonlinear regression program with the capability of calculating weighted least squares fits with iteratively recalculated weights can be used to fit probit models in the manner discussed in this section.

- Flexibility of models. Various functional forms such as probit, logit, or generalizations of these such as discussed by Prentice [26] can be specified. Transformations of concentration such as logarithm, square root, etc., can be specified. Alternatives to Abbott's correction for background can be specified (see e.g., Feder and Collins [1], Section XV), such as specifying background as an effective addition to the toxicant concentration. Centering and scale constants can be included in the models to reduce correlations among model terms and to improve the numerical convergence properties. Dose response curves resulting from different experiments can be compared with one another.

- Enhanced residual analysis capability. Predicted and residual values can be saved and further studied with subsequent statistical analyses and data displays.

We have fitted standard three parameter probit models with either logarithmic (common logs were used) or untransformed concentration levels. These models can be expressed as

$$p(conc)=p_0+(1-p_0)\Phi(\beta_0+\beta_1(z-m))$$

where p_0,p(conc) are the response rates at concentrations zero and "conc" respectively, z is either "conc" or $\log_{10}$(conc), m is a fixed centering constant to reduce the correlation among model terms and thereby improve convergence, $\Phi(\cdot)$ is the normal c.d.f., and p_0,β_0,β_1 are unknown parameters to be estimated from the model fit to the data.

An alternative adjustment for background mortality would be the model

$$p(conc)=\Phi(\alpha_0+\alpha_1 \log_{10}(conc+c))$$

where p(conc) is the response rate at once, c is the effective additive background concentration, and α_0,α_1,c are unknown constants to be estimated from the model fit. We decided not to use this alternative functional form because the shapes of the concentration-response relations displayed in Figures II.1-II.10 do not reflect this behavior. The wide range of concentrations at the low end of the mortality curve with nonzero but essentially constant mortality contradict the relationship that an additive background concentration would predict. We thus confined attention to Abbott's correction for background; however the alternative model would be no more difficult technically to specify and fit then the standard probit model.

Beaker-to-beaker heterogeneity within groups was accounted and adjusted for by reducing the actual sample sizes and numbers of responses to effective values, as discussed in Subsection VB, and carrying out subsequent analyses using these adjusted values. If desired, the t and F distributions with degrees of freedom based on those given in Tables V.3, V.4 might be substituted for the normal and chi square distributions when making inferences from the probit fits. However for LeBlanc's data sets these degrees of freedom are sufficiently large that the substitution would produce no differences of practical importance.

B. MAXIMUM LIKELIHOOD PROBIT ANALYSIS BY NONLINEAR LEAST SQUARES REGRESSION

Jennrich and Moore [25] show that for distributions in the exponential family, maximum likelihood estimation can be carried out by means of nonlinear least squares regression. This applies, in particular, to models based on the binomial distribution. Both BMDP [7] (P3R and PAR) and SAS [10] (PROC NLIN) contain nonlinear regression modules that can be used to fit various dose response curve models. Any nonlinear regression program with an iteratively reweighted least squares capability would suffice.

We fit the three parameter probit model discussed in the introduction, namely

$$p(conc)=p_0+(1-p_0)\Phi(\beta_0+\beta_1(z-m))$$

where z is conc or $\log_{10}(conc)$ and m is a fixed centering constant. The theory underlying the fit is discussed in general by Jennrich and Moore [25] and specifically for the probit model by Feder and Collins [1], Section XV. We discuss the details of the model fits and the resulting estimates below. We consider fits to the 21-day mortality responses from LeBlanc's Tests A and B and Goulden's Isophorone test. Under the assumptions of the nonlinear regression model, the dependent variable, X_i, has mean $\mu_i(\underset{\sim}{\theta})$ and variance $\sigma^2(\underset{\sim}{\theta})$, where $\theta\equiv(p_0,\beta_0,\beta_1)$, N_i is the effective sample size in the i-th group, and

$$\mu_i(\underset{\sim}{\theta})=p_i(\underset{\sim}{\theta})=p(conc_i)$$

$$\sigma_i^2(\underset{\sim}{\theta})=p_i(\underset{\sim}{\theta})(1-p_i(\underset{\sim}{\theta}))/N_i$$

We now consider the details of the fits to each of the data sets in turn.

LeBlanc Test A - Solvent Control Group - Logarithmic Concentration

We first discuss the details of the model fitting procedure using BMDPAR. (We used this program rather than BMDP3R because with BMDPAR we do not need to specify the functional forms of the derivatives of $\mu(\underset{\sim}{\theta})$). The BMDPAR program commands needed to generate the fit are given below. See the BMDP manual [7], pp 484-514 for further details. Various lines in the program command file are numbered. These numbered lines are explained further below.

Line 1 instructs the computer system to list the basic data prior to analysis.

Lines 2 attach the appropriate BMDP program and instruct it as to where the data are to be found. Note that the systems instructions above this point pertain only to CDC systems and undoubtedly differ at other installations.

Lines 3 are the basic input data. The four variables represent group number (one control and five treatment groups), concentration, effective sample size, and effective number of responses.

Lines 4 are a FORTRAN subroutine that specify the form of the dose response function and the form of the caseweights for the least squares fits. In this example, F is the probit dose response function, PHI is the standard normal c.d.f., and X(6) is the caseweight which is $1/\sigma_i^2(\underset{\sim}{\theta})$.

Lines 5 are standard BMDP control language commands which specify the source, form, identifiers, and desired transformations of the data. In

```
1
2
3
4
5
6
7
8
9
10
11
```

```
      140 REWIND,TAPE9.
      150 ATTACH,BMDPAR,ID=BMDP,PW=PDMB.
      160 FTN,B=LGO.
      170 BMDPAR(D=TAPE9)
      180 ◆EOR
      190 1.0         0.00063    59.3        2.2
      200 2.0         0.00294    59.3        7.4
      210 3.0         0.00286    59.3        5.1
      220 4.0         0.00741    59.3        5.3
      230 5.0         0.01509    5.9         2.9
      240 6.0         0.02890    59.3        58.5
      250 ◆EOR
      260       SUBROUTINE FUN(F,P,X,N,KASE,NVAR,NPAR,IPASS,XLOSS)
      270       DOUBLE PRECISION F,P,X,XLOSS
      280       DIMENSION P(NPAR),X(NVAR)
      290       IF(X(2) .LE. 0.00001) F=P(1)
      300       IF(X(2) .LE. 0.00001) GO TO 10
      310       Z=DLOG10(X(2))
      320       ARG=P(2)+P(3)◆(Z-X(8))
      330       PHI=(1+ERF(ARG/1.4142))/2
      340       F=P(1)+(1-P(1))◆PHI
      350    10 CONTINUE
      360       FF=F
      370       IF(F .LE. 0.001) FF=0.001
      380       IF(F .GE. 0.999) FF=0.999
      390       X(6)=X(3)/(FF◆(1.0-FF))
      400       RETURN
      410       END
      420 ◆EOR
      430 /PROBLEM TITLE IS "THREE PARAMETER PROBIT FIT USING BMDPAR NONL
      440                    REGRESSION PROGRAM".
      450 /INPUT     VARIABLE=4.
      460           FORMAT="(4F10.0)".
      470           UNIT=9.
      480 /VARIABLE NAME=GROUP,CONC,NTEST,NDEAD,P,CASEWT,LCONC,ZBAR.
      490            ADD=4.
      500 /TRANSFORM P=NDEAD/NTEST.
      510            CASEWT=1.0.
      520            LCONC=LOG(CONC).
      530            ZBAR=-2.113.
      540 /REGRESS   DEPENDENT=P.
      550            PARAMETER=3.
      560               TITLE="PROBIT FIT TO LOG10 CONC--LE BLANC TEST A
      570                     SOLVENT CONTOL GROUP".
      580            WEIGHT=CASEWT.
      590           HALVING=0.
      600            MEANSQUARE=1.0.
      610 /PARAMETER INITIAL=0.087476,-2.4432,3.107871.
      620            NAME=BKGD,INTCPT,SLOPE.
      630 /PLOT      VARIABLE=LCONC,CONC.
```

this example the centering constant, m, is denoted as ZBAR and is set equal to -2.113.

Lines 6 and 7 specify the dependent variable , P, and the weight variable, CASEWT. The observed P corresponds to the predicted F in the subroutine. The variable CASEWT is identified with X(6) in the subroutine. X(6) and F are evaluated for each case, for every iteration in the fitting procedure.

Line 8 is a technical fine point which instructs the fitting algorithm to find a stationary point rather to minimize the residual sum of squares. These two objectives differ because of the weights that depend on the unknown parameters. The distinction corresponds to the theoretical distinction between a maximum likelihood estimate and a minimum chi square estimate. While of theoretical interest, this detail is not of too much practical importance and so will not be pursued further. See the BMDP manual [7] or Jennrich and Moore [25] for further discussion.

Line 9 tells the program that the variances as specified in the model should be used for the calculation of the standard errors. More precisely, we specified caseweights equal to $1/\sigma_i^2(\theta)$, or equivalently variances of the dependent variable equal to $\sigma_i^2(\theta)$. However the least squares algorithm usually assumes that the variances are $k\sigma_i^2(\theta)$ where k is estimated from the residual mean square. The command in line 9 instructs the program to set k equal to 1.

Line 10 specifies the initial parameter values with which to start the iteration. These values are often estimated from a preliminary analysis or graphical display of the data. In this case they were based on the last set of parameter values from a previous run (not shown) which did not converge. The previous run was started with initial values based on graphing the data.

Line 11 specifies that observed, predicted, and residual values are to be plotted vs concentration and log concentration.

The output from these commands appears in Figures X.1 to X.4. Figure X.1, top, contains summary information about the problem specifications and input variables. The second portion of the figure contains the summary results of the iterative Gauss-Newton procedure. For each iteration the (weighted) residual sum of squares and updated parameter values are shown. The algorithm converges to a stationary point after 20 iterations. At this point, incremental changes in the residual sum of squares function are essentially zero. Thus this point corresponds to the maximum likelihood estimates. Note that this point does not minimize the residual sum of squares function, which was smaller at the starting values. This fine point was discussed in the introduction. The bottom portion of the figure contains statistics based on the model converged to. Parameter estimates are given for the three parameters in the model. Namely, $\hat{p}_0$=0.083, $\hat{\beta}_0$= -2.521, $\hat{\beta}_1$=8.108. The fitted model is

$$p(\text{conc})=\hat{p}_0+(1-\hat{p}_0)\Phi(\hat{\beta}_0+\hat{\beta}_1(\log_{10}(\text{conc})+2.113))$$

The estimated asymptotic standard errors of the parameter estimates are given beneath the estimates. These values are used for inferences.

Figure X.2 contains, for each case, the observed, predicted and residual values based on the fit. Also given are the estimated standard errors of the predictions and the values of the various input variables for that case. The plots in Figures X.3 and X.4 show good agreement between observed values and predictions based on the model. The greatest discrepancies occur at the control group and at the lowest concentration group, where the observed responses are most ragged. However even here, the discrepancies are not large.

The residual sum of squares is 3.01107 with 3 degrees of freedom. This value represents the chi square statistic for goodness of fit of the model. If the model fits the data then this statistic should have a chi square distribution with 3 d.f. Alternatively, using the residual degrees of freedom, 153, calculated from Table V.3 we might compare the residual mean square to the percentiles of the F-distribution with 3 and 153 degrees of freedom. The upper 90 percent point of the chi square distribution with 3 degrees of freedom is 6.25. There is thus no evidence of lack of fit.

An important purpose of fitting the probit model is to calculate point and confidence interval estimates of acceptable concentrations. Feder and Collins [1], Section XV, present a method of calculating approximate confidence intervals by means of Fieller's Theorem (Finney [4], pp 78-79). We present below an alternative method of calculating approximate confidence intervals by means of the delta method (Cramér [27], pp 366-367). We wish to calculate a point estimate and confidence interval on the concentration, C_L, such that $\Phi(\beta_0+\beta_1(\log_{10}C_L+2.113))=L$, where L is some specified incremental response rate over and above the control rate (e.g., L=0.05, 0.10, 0.25, etc.). L represents the response rate attributed to toxicant (over and above background level).

Let $z_L \equiv \log_{10}C_L+2.113$. We first calculate point and confidence interval estimates of z_L, and then translate them into corresponding estimates of C_L. The point estimate, $\hat{z}_L$, is

$$\hat{z}_L=(\Phi^{-1}(L)-\hat{\beta}_0)/\hat{\beta}_1 \equiv (f_L-\hat{\beta}_0)/\hat{\beta}_1.$$

Let $\hat{\ddagger}$ denote the estimated asymptotic variance-covariance matrix of $(\hat{\beta}_0,\hat{\beta}_1)$. Then

$$\text{Var}(\hat{z}_L)=(-1/\hat{\beta}_1,(\hat{\beta}_0-f_L)/\hat{\beta}_1^2)\ \hat{\ddagger}\ (-1/\hat{\beta}_1,(\hat{\beta}_0-f_L)/\hat{\beta}_1^2)'$$

An approximate $1-\alpha$ confidence interval interval on z_L is

$$\hat{z}_L \pm \xi_{\alpha/2}[\hat{\text{Var}}(\hat{z}_L)]^{1/2}=(\ell,u)$$

where $\xi_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution. The theoretical justification of these expressions is given in Appendix AX.

The values of $\hat{\beta}_0, \hat{\beta}_1, \hat{\ddagger}$ are obtained from the BMDPAR output, in particular from Figure X.1. The corresponding estimate and confidence interval for C_L is

$$\hat{C}_L = 10^{\hat{z}_L - 2.113} \text{ and } C_L \in (10^{\ell - 2.113}, 10^{u - 2.113})$$

For the LeBlanc Test A solvent control data, $\hat{\beta}_0 = -2.52089, \hat{\beta}_1 = 8.27337$

$$\hat{\ddagger} = \begin{pmatrix} 1.20098 & 0 \\ 0 & 2.49065 \end{pmatrix} \begin{pmatrix} 1.0000 & -0.9546 \\ -0.9546 & 1.0000 \end{pmatrix} \begin{pmatrix} 1.20098 & 0 \\ 0 & 2.49065 \end{pmatrix} = \begin{pmatrix} 1.44235 & -2.8554 \\ -2.8554 & 6.2033 \end{pmatrix}$$

The results of the calculations are given below.

LeBlanc Test A - Solvent Control Group-Point Estimates and 95 Percent Confidence Intervals on Various Percentiles of the Probit Fit--by Delta Method

L	$f_L \equiv \Phi^{-1}(L)$	$\hat{Z}_L$	ℓ	u	$\hat{C}_L$	lower conf bound	upper conf bound
0.05	-1.645	0.1059	-0.1198	0.3315	0.0098	0.0059	0.0165
0.10	-1.282	0.1497	-0.0522	0.3516	0.0109	0.0068	0.0173
0.25	-0.674	0.2232	0.0597	0.3868	0.0129	0.0088	0.0188
0.35	-0.385	0.2582	0.1118	0.4045	0.0140	0.0099	0.0196
0.50	0.000	0.3047	0.1797	0.4297	0.0155	0.0117	0.0207

These results show that the dose response curve rises very rapidly between the estimated 5th and 50th percentiles. In fact the upper confidence bound on $C_{.05}$ exceeds the estimate of C_{50}. Thus dose response curve perecentiles with very different biological implications cannot be well separated based on the results of this test.

We carried out similar analyses using the water control group rather than the solvent control group. The three parameter probit model fitted to logarithmic concentration again fitted the data well. The details of the fit are not shown, however the estimated background mortality rate is 0.117 (with a standard error of 0.021) as compared to 0.0831 (with a standard error of 0.018) based on the solvent control group. These adjusted background mortality rates are not significantly different (statistically or biologically).

The results of the percentile estimates and confidence interval calculations are given below.

LeBlanc Test A - Water Control Group-Point Estimates and 95 Percent Confidence Intervals on Various Percentiles of Probit Fit

L	$f_L \equiv \Phi^{-1}(L)$	$\hat{C}_L$	lower conf bound	upper conf bound
0.05	-1.645	0.0103	0.0062	0.0172
0.10	-1.282	0.0114	0.0072	0.0180
0.25	-0.674	0.0134	0.0092	0.0194
0.35	-0.385	0.0144	0.0104	0.0201
0.50	0.000	0.0160	0.0120	0.0212

The results are the same, for practical considerations as those based on the solvent control group. Thus even though the unadjusted mortality rates in the water and solvent control groups are (statistically) significantly different, the estimates of dose response curve percentiles are virtually unaffected.

LeBlanc Test B--Combined Control Groups

We now consider the results of fitting the three parameter probit model to the 21-day mortality results from LeBlanc's Test B. The results in the water and solvent control groups were combined and used for comparison purposes. The measured concentration levels in these two groups were averaged. Effective sample sizes and numbers of responses, as shown in Table V.4, were used in the analysis.

We first attempted to fit a probit model using logarithmic concentration. The results are shown in Figures X.5-X.7. The algorithm would not converge when the increment halving option was removed. Thus the values converged to are not strictly maximum likelihood estimates. However the sum of squares is nearly at a stationary point and the parameter estimates have settled down. The plot in Figure X.7 shows about as good a fit as can be expected to such nonmonotone responses. The residual sum of squares, 5.7839 with 3 degrees of freedom, is large but not statistically significant. (The upper 90 percent point of the chi square distribution with 3 degrees of freedom is 6.251.) Thus we will use these estimates. The parameter estimates in this model are so highly intercorrelated that standard error and correlation estimates cannot be properly calculated. This is the reason for the message at the top of Figure X.6. The reason for this intercorrelation is clearly seen in Figure X.7. The response curve is flat throughout

most of the range of concentrations. Only the highest treatment group has a mortality rate substantially in excess of the control rate, and the effective sample size in that group is just 8.96. Thus there is very poor information about the slope of the response curve.

We next refitted the probit model using the same data but with untransformed concentration. In particular, the model fitted was

$$p(conc)=p_0+(1-p_0)\Phi(\beta_0+\beta_1(conc-0.0250))$$

The results are shown in Figures X.8-X.10. As in the first attempt, the algorithm would not converge when the increment halving option was removed. This option was therefore retained and so the values converged to are again not strictly maximum likelihood estimates. However, as the iteration process has pretty much settled down and since the fit appears to be good (see Figure X.10), we will use these estimates. The residual sum of squares, 5.8748 with 3 degrees of freedom, is again large but not statistically significant. The intercorrelation among the estimates has been reduced to the extent that estimated correlations and standard errors can at least be calculated. Note that the estimated slope is 15.9695 with a standard error of 12.9349. Thus $\hat{\beta}_1$ is not significantly different from 0. This high standard error may be due to the very high intercorrelation between the slope and intercept estimates. This in turn is due to the very limited information about the slope. The mortality rate is essentially constant throughout most of the range of concentrations.

The parameters from the model fit were used to calculate point estimates of and confidence intervals on the dose response curve percentiles. The results of these calculations are given below.

LeBlanc Test B - Combined Control Groups-Point Estimates and 95 Percent Confidence Intervals on Various Percentiles of Probit Fit

L	$f_L \equiv \Phi^{-1}(L)$	$\hat{C}_L$	lower conf bound	upper conf bound
0.05	-1.645	0.1332	-0.0435	0.3099
0.10	-1.282	0.1559	0.0130	0.2989
0.25	-0.674	0.1940	0.1036	0.2844
0.35	-0.385	0.2121	0.1420	0.2822
0.50	0.000	0.2362	0.1803	0.2921

The confidence intervals, particularly those corresponding to the lower percentiles, are so wide as to be useless. This is undoubtedly due to the high degree of uncertainty in the parameter estimates, as indicated by their very large standard errors. We thus conclude that we cannot make very strong inferences, based on the results of this test, about the concentrations corresponding to various levels of mortality.

The final example pertains to the 21-day mortality data from Goulden's test on isophorone.

Goulden Isophorone Test--Logarithmic Concentration

As there was just one control group in this test, there is no question about which control group or combination of control groups to use for comparison purposes. The dose response curve is estimated from the results only on the multiple housed daphnids. As there is no evidence of beaker-to-beaker heterogeneity within groups, there is no need to adjust the sample sizes and numbers of responses. The model fitted was

$$p(\text{conc}) = p_0 + (1-p_0)\Phi(\beta_0 + \beta_1(\log_{10}(\text{conc}) - 1.8352))$$

Let z_L denote $\log_{10}C_L - 1.8352$.

The results from the probit fit are shown in Figures X.11-X.13. The Gauss-Newton algorithm converged to the maximum likelihood estimates shown in iteration 15 in Figure XI.11. The residual sum of squares, 0.6486 with 3 degrees of freedom, is very small. The fit looks quite good in Figure X.13.

The parameters from the model fit were used to calculate point estimates of and confidence intervals on the dose response curve percentiles. The results of the calculations are given below.

Goulden Isophorone Test-Point Estimates and 95 Percent Confidence Intervals on Various Percentiles of Probit Fit

L	$f_L \equiv \Phi^{-1}(L)$	$\hat{C}_L$	lower conf bound	upper conf bound
0.05	-1.645	89.20	59.21	134.38
0.10	-1.282	99.71	70.76	140.51
0.25	-0.674	120.17	94.55	152.74
0.35	-0.385	131.32	107.65	160.19
0.50	0.000	147.80	125.75	173.70

These confidence intervals are fairly wide, probably due to the small sample sizes in the test. Greater precision in the estimation of these percentiles must await the results of more extensive tests.

REGRESSION TITLE
PROBIT FIT TO LOG10 CONC--LE BLANC TEST A- SOLVENT CONTOL GROUP

REGRESSION NUMBER	0
DEPENDENT VARIABLE	P
WEIGHTING VARIABLE	CASEWT
NUMBER OF PARAMETERS	3
NUMBER OF CONSTRAINTS	0
TOLERANCE FOR PIVOTING	1.0E-08
TOLERANCE FOR CONVERGENCE	1.0E-05
MAXIMUM NUMBER OF ITERATIONS	50
MAXIMUM NUMBER OF INCREMENT HALVINGS	0
SPECIFIED RESIDUAL MEAN SQUARE	1.000

PARAMETERS TO BE ESTIMATED

	1 BKGD	2 INTCPT	3 SLOPE
MINIMUM	**********	**********	**********
MAXIMUM	**********	**********	**********
INITIAL	.087476	-2.443200	8.107871

USING THE ABOVE SPECIFICATIONS THIS PROGRAM COULD USE UP TO 356 CASES.

NUMBER OF CASES READ. 6

VARIABLE NO.	NAME	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
1	GROUP	3.500000	1.870829	1.000000	6.000000
2	CONC	.009638	.010749	.000630	.028900
3	NTEST	50.300000	21.800459	5.800000	59.200000
4	NDEAD	13.550000	22.098394	2.200000	58.500000
5	P	.304054	.375218	.037162	.988176
7	LCONC	-2.294423	.593459	-3.200659	-1.539102
8	ZBAR	-2.113000	0.000000	-2.113000	-2.113000
6	CASEWT	NOT COMPUTED		1.000000	1.000000

ITER. NO.	INCR. HALV.	RESIDUAL SUM OF SQUARES	PARAMETERS 1 BKGD	2 INTCPT	3 SLOPE
0	0	4.416431	.087476	-2.443200	8.918658
0	0	3.354416	.087476	-2.687520	8.107871
0	0	3.124958	.096224	-2.443200	8.107871
0	0	2.941857	.087476	-2.443200	8.107871
1	0	3.254782	.082424	-2.258778	7.439695
2	0	3.011160	.083130	-2.513760	8.253810
3	0	3.011170	.083127	-2.518516	8.267223
4	0	3.010068	.083163	-2.531238	8.297570
5	0	3.013996	.083043	-2.525964	8.286965
6	0	3.011060	.083130	-2.521294	8.273944
7	0	3.022463	.082829	-2.555351	8.350336
8	0	3.011103	.083129	-2.521572	8.274304
9	0	3.010089	.083160	-2.522638	8.277251
10	0	3.011214	.083126	-2.521101	8.273989
11	0	3.011165	.083127	-2.521233	8.274121
12	0	3.011124	.083129	-2.521105	8.273836

Figure X.1. Output from BMDPAR applied to LeBlanc test A 21 day mortality data--solvent control group--logarithmic concenctration.

13	0	3.011061	.083130	-2.520910	8.273407
14	0	3.010966	.083133	-2.521012	8.273684
15	0	3.011097	.083129	-2.520824	8.273304
16	0	3.011070	.083130	-2.520896	8.273375
17	0	3.011069	.083130	-2.520892	8.273365
18	0	3.011069	.083130	-2.520893	8.273369
19	0	3.011070	.083130	-2.520893	8.273368
20	0	3.011069	.083130	-2.520893	8.273369

PROBIT FIT TO LOG10 CONC--LE BLANC TEST A- SOLVENT CONTOL GROUP

THE RESIDUAL SUM OF SQUARES (= 2.94186) WAS SMALLEST WITH THE FOLLOWING PARAMETER VALUES

1 BKGD	2 INTCPT	3 SLOPE
8.747600E-02	-2.44320	8.10787

THE FINAL VALUE (= 3.01107) WILL BE USED IN FUTURE COMPUTATIONS.
CORRESPONDING PARAMETER VALUES ARE

1 BKGD	2 INTCPT	3 SLOPE
8.313004E-02	-2.52089	8.27337

ESTIMATE OF ASYMPTOTIC CORRELATION MATRIX

		BKGD 1	INTCPT 2	SLOPE 3
BKGD	1	1.0000		
INTCPT	2	-.2071	1.0000	
SLOPE	3	.1886	-.9546	1.0000

THE ESTIMATED MEAN SQUARE ERROR IS .9806
THE SPECIFIED VALUE OF THE MEAN SQUARE ERROR (1.000)NOT THE COMPUTED VALUE IS USED IN COMPUTING STANDARD DEVIATIONS

ESTIMATES OF ASYMPTOTIC STANDARD DEVIATIONS OF PARAMETER ESTIMATES WITH 3 DEGREES OF FREEDOM ARE

1 BKGD	2 INTCPT	3 SLOPE
1.842787E-02	1.20098	2.49065

Figure X.1. Continued.

PROBIT FIT TO LOG10 CONC--LE BLANC TEST A- SOLVENT CONTOL GROUP

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	1 GROUP	2 CONC	3 NTEST	4 NDEAD
1		-.045968	.037162	.083130	.018428	1.000000	.000630	59.200000	2.200000
2		.041870	.125000	.083130	.018428	2.000000	.002940	59.200000	7.400000
3		.003019	.086149	.083130	.018428	3.000000	.002860	59.200000	5.100000
4		.001158	.087838	.086680	.020227	4.000000	.007410	59.200000	5.200000
5		-.002268	.500000	.502268	.198854	5.000000	.015090	5.800000	2.900000
6		.000065	.988176	.988111	.013983	6.000000	.028900	59.200000	58.500000

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	6 CASEWT	7 LCONC	8 ZBAR
1		-.045968	.037162	.083130	.018428	776.704761	-3.200659	-2.113000
2		.041870	.125000	.083130	.018428	776.704752	-2.531653	-2.113000
3		.003019	.086149	.083130	.018428	776.704756	-2.543634	-2.113000
4		.001158	.087838	.086680	.020227	747.792155	-2.130182	-2.113000
5		-.002268	.500000	.502268	.198854	23.200477	-1.821311	-2.113000
6		.000065	.988176	.988111	.013983	5039.257109	-1.539102	-2.113000

Figure X.2. Observed, predicted and residual values from probit fit to LeBlanc test A 21 day mortality data--solvent control group--logarthmic concentration

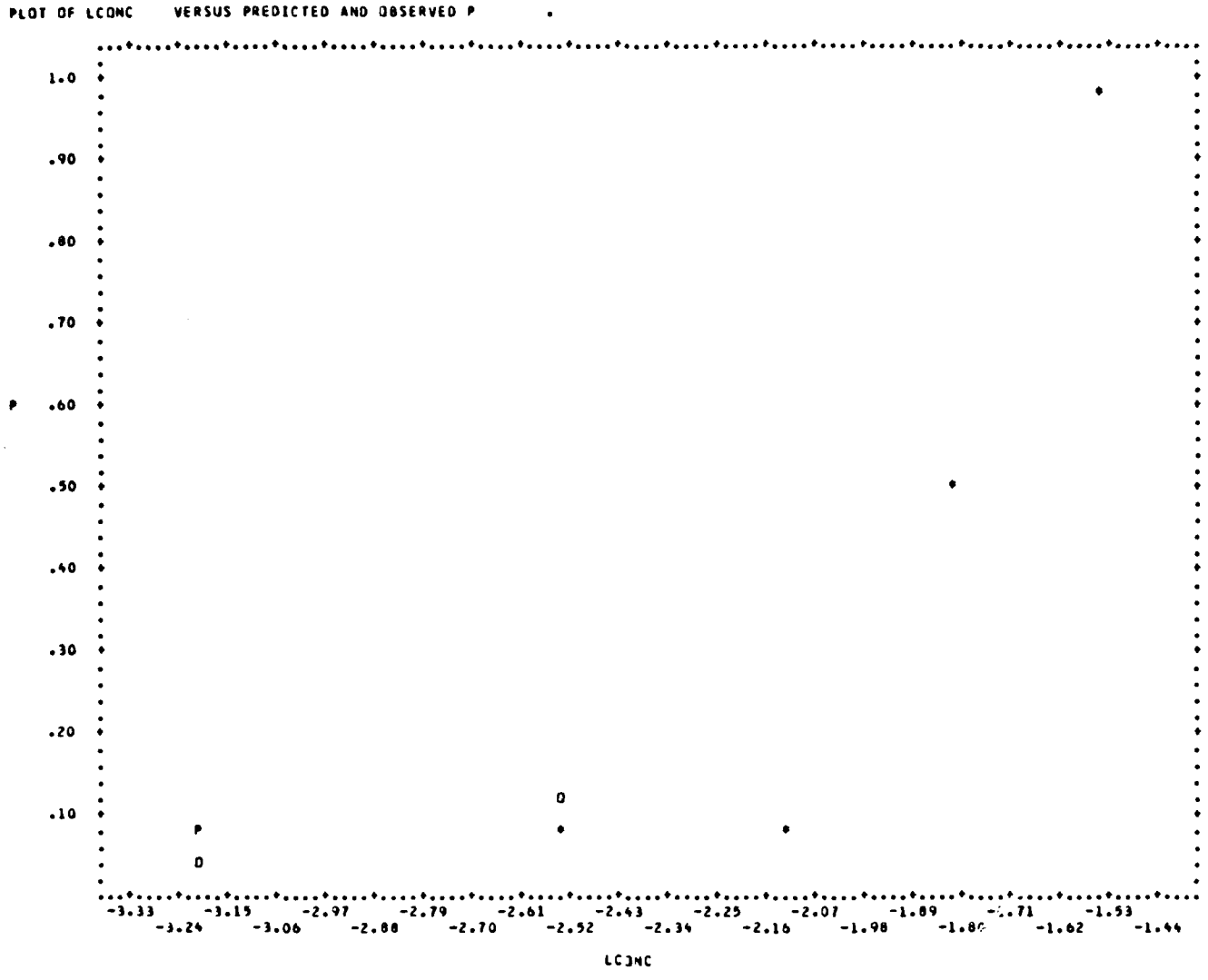


Figure X.3. Observed and predicted values from fit in Figure X.1 versus logarithmic concentration.

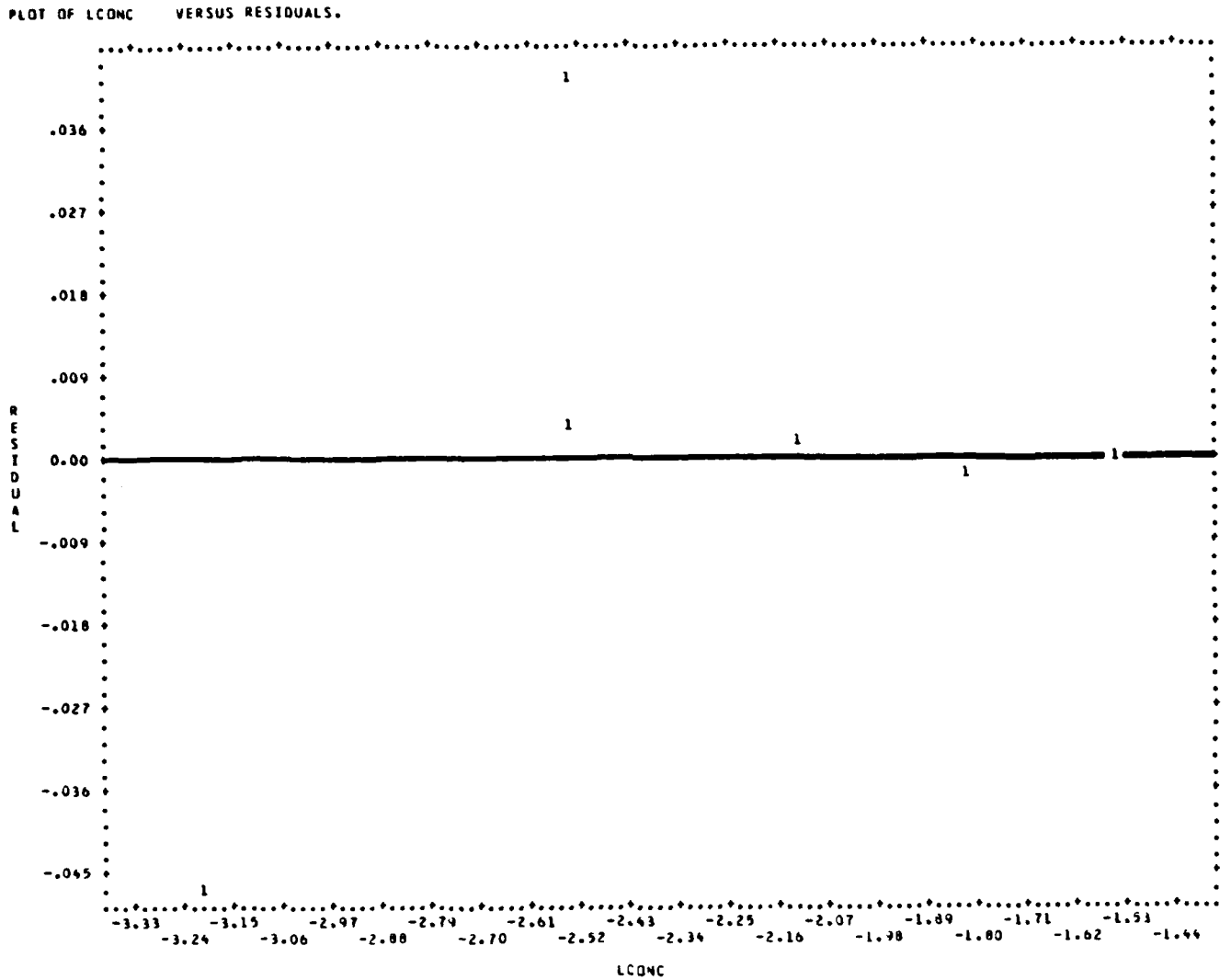


Figure X.4. Residuals from fit in Figure X.1 versus logarithmic concentration.

REGRESSION TITLE
PROBIT FIT TO LOG10 CONC--LE BLANC TEST B- COMBINED CONTOL GROUPS

REGRESSION NUMBER	0
DEPENDENT VARIABLE	P
WEIGHTING VARIABLE	CASEWT
NUMBER OF PARAMETERS	3
NUMBER OF CONSTRAINTS	0
TOLERANCE FOR PIVOTING	1.0E-08
TOLERANCE FOR CONVERGENCE	1.0E-05
MAXIMUM NUMBER OF ITERATIONS	50
MAXIMUM NUMBER OF INCREMENT HALVINGS	10
SPECIFIED RESIDUAL MEAN SQUARE	1.000

PARAMETERS TO BE ESTIMATED

	1 BKGD	2 INTCPT	3 SLOPE
MINIMUM	**********	**********	**********
MAXIMUM	**********	**********	**********
INITIAL	.111563	-369.094320	185.460983

USING THE ABOVE SPECIFICATIONS THIS PROGRAM COULD USE UP TO 356 CASES.

NUMBER OF CASES READ. 6

VARIABLE NO.	NAME	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
1	GROUP	3.500000	1.870829	1.000000	6.000000
2	CONC	.072666	.091030	.000250	.246300
3	NTEST	59.033333	33.675597	8.960000	115.200000
4	NDEAD	7.033000	4.564751	2.800000	15.100000
5	P	.188489	.218590	.049611	.626116
7	LCONC	-1.648657	1.046670	-3.602060	-.638536
8	ZBAR	-2.600000	0.000000	-2.600000	-2.600000
6	CASEWT	NOT COMPUTED		1.000000	1.000000

ITER. NO.	INCR. HALV.	RESIDUAL SUM OF SQUARES	PARAMETERS 1 BKGD	2 INTCPT	3 SLOPE
0	0	1253.550889	.111563	-369.094320	204.007081
0	0	23.720308	.111563	-406.003752	185.460983
0	0	6.147517	.122719	-369.094320	185.460983
0	0	5.795342	.111563	-369.094320	185.460983
1	5	5.787135	.111386	-369.094320	185.439586
2	6	5.786767	.111472	-369.103974	185.439753
3	2	5.784699	.112864	-369.104611	185.438353
4	3	5.784375	.111994	-369.104207	185.439242
5	4	5.783359	.112375	-369.104384	185.438853
6	10	5.783059	.112381	-369.104387	185.438847
7	10	5.783960	.112387	-369.104390	185.438840
8	9	5.783859	.112375	-369.104384	185.438853
9	10	5.783359	.112381	-369.104387	185.438847
10	10	5.783960	.112387	-369.104390	185.438840
11	9	5.783859	.112375	-369.104358	185.438840

Figure X.5. Output from BMDPAR applied to LeBlanc test B 21 day mortality data--combined control group--logarithmic concentration.

PROBIT FIT TO LOG10 CONC--LE BLANC TEST 9- COMBINED CONTOL GROUPS

THE RESIDUAL SUM OF SQUARES (= 5.78356) WAS SMALLEST WITH THE FOLLOWING PARAMETER VALUES

1 BKGD	2 INTCPT	3 SLOPE
.112375	-369.104	185.439

* * * HIGH CORRELATIONS AMONG THE PARAMETER ESTIMATES PREVENTS COMPUTATION OF THE COMPLETE COVARIANCE MATRIX
THE FOLLOWING CORRELATIONS AND STANDARD DEVIATIONS ARE CONDITIONAL UPON SOME OF THE PARAMETERS BEING HELD CONSTANT.
SOME OF THE UNCONDITIONAL VALUES (IF THEY COULD BE COMPUTED) MAY BE MUCH LARGER.

ESTIMATE OF ASYMPTOTIC CORRELATION MATRIX

		BKGD 1	INTCPT 2	SLOPE 3
BKGD	1	1.0000		
INTCPT	2	-.6445	1.0000	
SLOPE	3	0.0000	0.0000	0.0000

THE ESTIMATED MEAN SQUARE ERROR IS 1.446
THE SPECIFIED VALUE OF THE MEAN SQUARE ERROR (1.000)NOT THE COMPUTED VALUE IS USED IN COMPUTING STANDARD DEVIATIONS

ESTIMATES OF ASYMPTOTIC STANDARD DEVIATIONS OF PARAMETER ESTIMATES WITH 4 DEGREES OF FREEDOM ARE

1 BKGD	2 INTCPT	3 SLOPE
1.698961E-02	.466468	0.

Figure X.6. Summary of results from BMDPAR probit fit to LeBlanc test B 21 day mortality data--combined control group--logarithmic concentration.

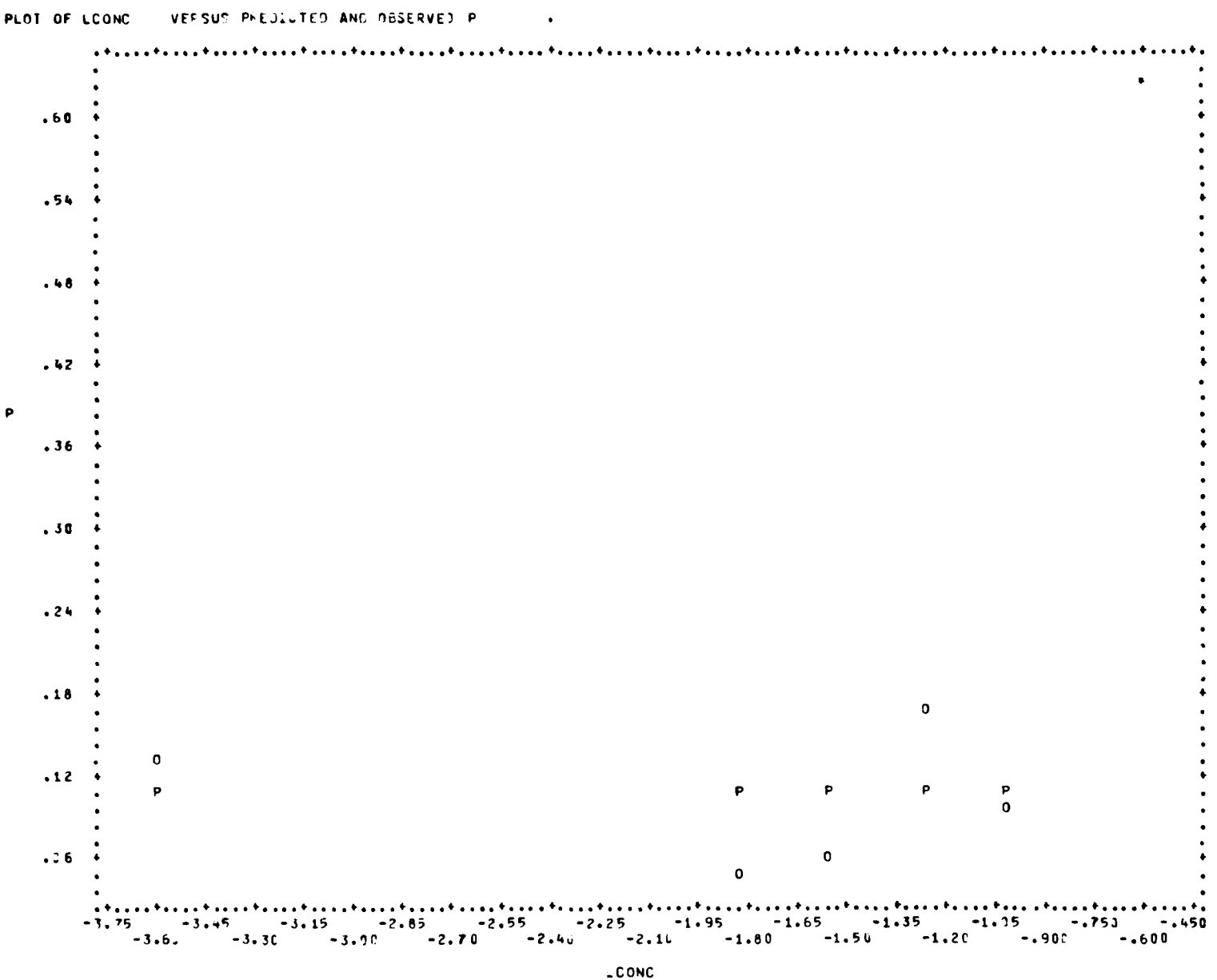


Figure X.7. Observed and predicted values from fit in Figures X.5 and X.6 versus logarithmic concentration.

REGRESSION TITLE
PROBIT FIT TO UNTNSFMD CONC--LE BLANC TEST B- COMBINED CONTOL GROUPS

REGRESSION NUMBER	0
DEPENDENT VARIABLE	P
WEIGHTING VARIABLE	CASEWT
NUMBER OF PARAMETERS	3
NUMBER OF CONSTRAINTS	0
TOLERANCE FOR PIVOTING	1.0E-08
TOLERANCE FOR CONVERGENCE	1.0E-05
MAXIMUM NUMBER OF ITERATIONS	50
MAXIMUM NUMBER OF INCREMENT HALVINGS	10
SPECIFIED RESIDUAL MEAN SQUARE	1.000

PARAMETERS TO BE ESTIMATED

	1 BKGD	2 INTCPT	3 SLOPE
MINIMUM	**********	**********	**********
MAXIMUM	**********	**********	**********
INITIAL	.109264	-3.377945	15.997613

USING THE ABOVE SPECIFICATIONS THIS PROGRAM COULD USE UP TO 356 CASES.

NUMBER OF CASES READ. 6

VARIABLE NO.	NAME	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
1	GROUP	3.500000	1.870829	1.000000	6.000000
2	CONC	.072888	.091030	.000250	.246300
3	NTEST	59.093333	33.675597	8.960000	115.200000
4	NDEAD	7.035000	4.564751	2.800000	15.100000
5	P	.188409	.218590	.048611	.626116
7	LCONC	-1.648057	1.046670	-3.602060	-.608536
8	ZBAR	.025000	0.000000	.025000	.025000
6	CASEWT	NOT COMPUTED		1.000000	1.000000

ITER. NO.	INCR. HALV.	RESIDUAL SUM OF SQUARES	PARAMETERS 1 BKGD	2 INTCPT	3 SLOPE
0	0	6.472368	.109264	-3.715740	15.997613
0	0	6.408134	.109264	-3.377945	17.597374
0	0	6.164202	.120190	-3.377945	15.997613
0	0	5.875631	.109264	-3.377945	15.997613
1	6	5.875580	.109338	-3.373380	15.972977
2	6	5.875557	.109406	-3.369084	15.950784
3	8	5.875554	.109423	-3.368056	15.945383
4	5	5.875460	.109291	-3.377682	15.995102
5	6	5.875405	.109357	-3.373401	15.972973
6	6	5.875373	.109425	-3.368869	15.950095
7	6	5.875385	.109496	-3.364923	15.929198
8	5	5.875095	.109364	-3.375177	15.982158
9	6	5.875067	.109432	-3.370998	15.960444
10	10	5.875067	.109436	-3.370721	15.959033
11	8	5.875063	.109454	-3.369717	15.953748
12	5	5.874946	.109321	-3.379323	14.003532

Figure X.8. Output from BMDPAR applied to LeBlanc test B 21 day mortality data--combined control groups--untransformed concentration.

13	6	5.874897	.109388	-3.375115	15.981613
14	7	5.874488	.109354	-3.377967	15.995237
15	6	5.874815	.109423	-3.373985	15.974412
16	10	5.874815	.109427	-3.373717	15.973013
17	6	5.874792	.109498	-3.369703	15.952042
18	7	5.874771	.109459	-3.371734	15.963582
19	7	5.874761	.109424	-3.373929	15.974978
20	9	5.874759	.109442	-3.372871	15.969469

PROBIT FIT TO UNTNSFMD CONC--LE BLANC TEST B- COMBINED CONTOL GROUPS

THE RESIDUAL SUM OF SQUARES (= 5.87476) WAS SMALLEST WITH THE FOLLOWING PARAMETER VALUES

1 BKGD	2 INTCPT	3 SLOPE
.109442	-3.37287	15.9695

ESTIMATE OF ASYMPTOTIC CORRELATION MATRIX

		BKGD 1	INTCPT 2	SLOPE 3
BKGD	1	1.0000		
INTCPT	2	-.5474	1.0000	
SLOPE	3	.5291	-.9868	1.0000

THE ESTIMATED MEAN SQUARE ERROR IS 1.958
THE SPECIFIED VALUE OF THE MEAN SQUARE ERROR (1.000)NOT THE COMPUTED VALUE IS USED IN COMPUTING STANDARD DEVIATIONS

ESTIMATES OF ASYMPTOTIC STANDARD DEVIATIONS OF PARAMETER ESTIMATES WITH 3 DEGREES OF FREEDOM ARE

1 BKGD	2 INTCPT	3 SLOPE
2.022659E-02	2.80322	12.9349

Figure X.8. Continued.

PROBIT FIT TO UNTNSFMD CONC--LE BLANC TEST B- COMBINED CONTOL GROUPS

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	1 GROUP	2 CONC	3 NTEST	4 NDEAD
1		.021562	.131076	.109515	.019776	1.000000	.000250	115.200000	15.100000
2		-.061007	.048611	.109618	.019231	2.000000	.014380	57.600000	2.800000
3		-.047347	.062500	.109847	.018434	3.000000	.028500	57.600000	3.600000
4		.052023	.163194	.111171	.017080	4.000000	.055400	57.600000	9.400000
5		-.020162	.098958	.119120	.041560	5.000000	.092500	57.600000	5.700000
6		.014380	.626116	.611736	.162799	6.000000	.246300	8.960000	5.610000

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	6 CASEWT	7 LCONC	8 ZBAR
1		.021562	.131076	.109515	.019776	1181.280873	-3.602060	.025000
2		-.061007	.048611	.109618	.019231	590.152504	-1.842241	.025000
3		-.047347	.062500	.109847	.018434	589.075322	-1.545155	.025000
4		.052023	.163194	.111171	.017080	582.925076	-1.256490	.025000
5		-.020162	.098958	.119120	.041560	548.935148	-1.033858	.025000
6		.014380	.626116	.611736	.162799	37.723424	-.608536	.025000

Figure X.9. Observed, predicted and residual values from probit fit to LeBlanc test B 21 day mortality data--combined control groups--untransformed concentration.

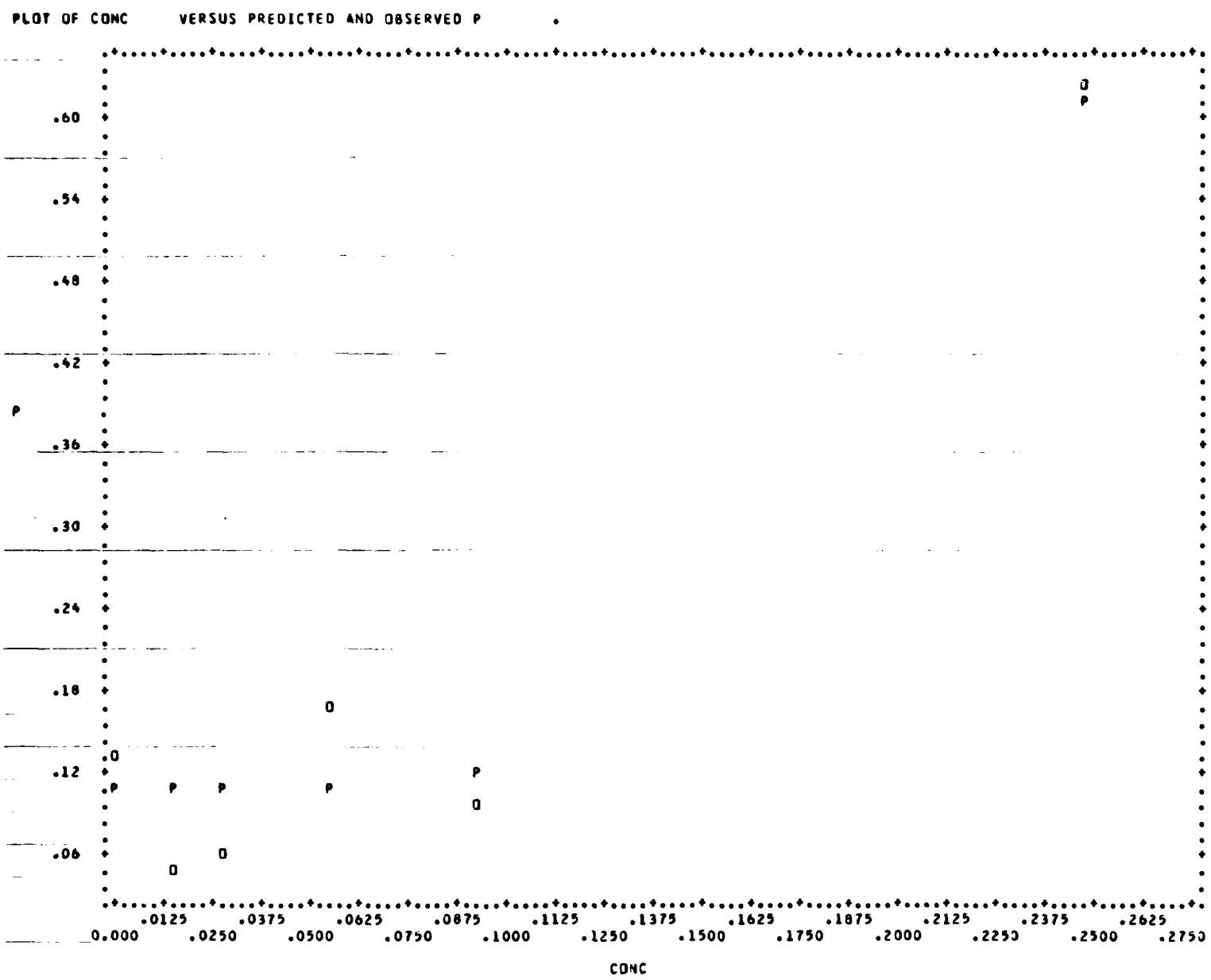


Figure X.10. Observed and predicted values from fit in Figure X.8 versus concentration.

REGRESSION TITLE
PROBIT FIT TO LOG10 CONC--GOULDEN ISOPHORONE TEST

REGRESSION NUMBER	0
DEPENDENT VARIABLE	P
WEIGHTING VARIABLE	CASEWT
NUMBER OF PARAMETERS	3
NUMBER OF CONSTRAINTS	0
TOLERANCE FOR PIVOTING	1.0E-08
TOLERANCE FOR CONVERGENCE	1.0E-05
MAXIMUM NUMBER OF ITERATIONS	50
MAXIMUM NUMBER OF INCREMENT HALVINGS	0
SPECIFIED RESIDUAL MEAN SQUARE	1.000

PARAMETERS TO BE ESTIMATED

	1 BKGD	2 INTCPT	3 SLOPE
MINIMUM	**********	**********	**********
MAXIMUM	**********	**********	**********
INITIAL	.095826	-2.534960	7.557686

USING THE ABOVE SPECIFICATIONS THIS PROGRAM COULD USE UP TO 355 CASES.

NUMBER OF CASES READ. 6

VARIABLE NO.	NAME	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
1	GROUP	3.500000	1.870829	1.000000	6.000000
2	CONC	*85.000000*	*79.686887*	*0.000000*	*200.000000*
3	NTEST	15.000000	0.000000	15.000000	15.000000
4	NDEAD	4.666667	4.844241	1.000000	13.000000
5	P	.311111	.322949	.066667	.866667
7	LCONC	1.539243	.878025	0.000000	2.303196
8	ZBAR	1.835200	0.000000	1.835200	1.835200
6	CASEWT	NOT COMPUTED		1.000000	1.000000

ITER. NO.	INCR. HALV.	RESIDUAL SUM OF SQUARES	PARAMETERS 1 BKGD	2 INTCPT	3 SLOPE
0	0	2.040613	.095826	-2.534960	8.313455
0	0	1.570471	.095826	-2.788456	7.557686
0	0	.694992	.105409	-2.534960	7.557686
0	0	.636014	.095826	-2.534960	7.557686
1	0	.871520	.088433	-2.144896	6.392476
2	0	.649436	.090509	-2.515308	7.503391
3	0	.648224	.090633	-2.502199	7.482587
4	0	.648524	.090596	-2.510009	7.504912
5	0	.645323	.091215	-2.508221	7.495072
6	0	.648569	.090582	-2.508932	7.501161
7	0	.648605	.090576	-2.508678	7.500956
8	0	.648599	.090578	-2.508923	7.501410
9	0	.648603	.090577	-2.508852	7.501216
10	0	.648585	.090580	-2.508891	7.501318
11	0	.648588	.090579	-2.508891	7.501323
12	0	.648587	.090580	-2.508895	7.501323

Figure X.11. Output from BMDPAR applied to Goulden isophorone test 21 day mortality data--logarithmic concentration.

13	0	.648587	.090580	-2.508893	7.501321
14	0	.648587	.090580	-2.508895	7.501325
15	0	.648587	.090580	-2.508895	7.501324

PROBIT FIT TO LOG10 CONC--GOULDEN ISOPHORONE TEST

THE RESIDUAL SUM OF SQUARES (= .636014) WAS SMALLEST WITH THE FOLLOWING PARAMETER VALUES

1 BKGD	2 INTCPT	3 SLOPE
9.582600E-02	-2.53496	7.55769

THE FINAL VALUE (= .648587) WILL BE USED IN FUTURE COMPUTATIONS.
CORRESPONDING PARAMETER VALUES ARE

1 BKGD	2 INTCPT	3 SLOPE
9.057960E-02	-2.50889	7.50132

ESTIMATE OF ASYMPTOTIC CORRELATION MATRIX

		BKGD 1	INTCPT 2	SLOPE 3
BKGD	1	1.0000		
INTCPT	2	-.3238	1.0000	
SLOPE	3	.2590	-.9633	1.0000

THE ESTIMATED MEAN SQUARE ERROR IS .2120
THE SPECIFIED VALUE OF THE MEAN SQUARE ERROR (1.000)NOT THE COMPUTED VALUE IS USED IN COMPUTING STANDARD DEVIATIONS

ESTIMATES OF ASYMPTOTIC STANDARD DEVIATIONS OF PARAMETER ESTIMATES WITH 3 DEGREES OF FREEDOM ARE

1 BKGD	2 INTCPT	3 SLOPE
4.258345E-02	.955959	2.51776

Figure X.11. Continued.

PROBIT FIT TO LOG10 CONC--GOULDEN ISOPHORONE TEST

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	1 GROUP	2 CONC	3 NTEST	4 NDEAD
1		.042754	.133333	.090580	.042583	1.000000	0.000000	15.000000	2.000000
2		-.023913	.066667	.090580	.042583	2.000000	10.000000	15.000000	1.000000
3		-.024101	.066667	.090768	.042300	3.000000	50.000000	15.000000	1.000000
4		.017056	.200000	.182944	.085746	4.000000	100.000000	15.000000	3.000000
5		-.029452	.533333	.562785	.090979	5.000000	150.000000	15.000000	8.000000
6		.014175	.866667	.852492	.080447	6.000000	200.000000	15.000000	13.000000

CASE NO.	NAME	RESIDUAL	OBSERVED 5 P	PREDICTED 5 P	STD. DEV. PREDICTED	6 CASEWT	7 LCONC	8 ZBAR
1		.042754	.133333	.090580	.042583	182.094218	0.000000	1.835200
2		-.023913	.066667	.090580	.042583	182.094218	1.041393	1.835200
3		-.024101	.066667	.090768	.042300	181.753976	1.707570	1.835200
4		.017056	.200000	.182944	.085746	100.350801	2.004321	1.835200
5		-.029452	.533333	.562785	.090979	60.961239	2.178977	1.835200
6		.014175	.866667	.852492	.080447	119.284944	2.303196	1.835200

Figure X.12. Observed, predicted and residual values from probit fit to Goulden isophorone test 21 day mortality data--logarithmic concentration.

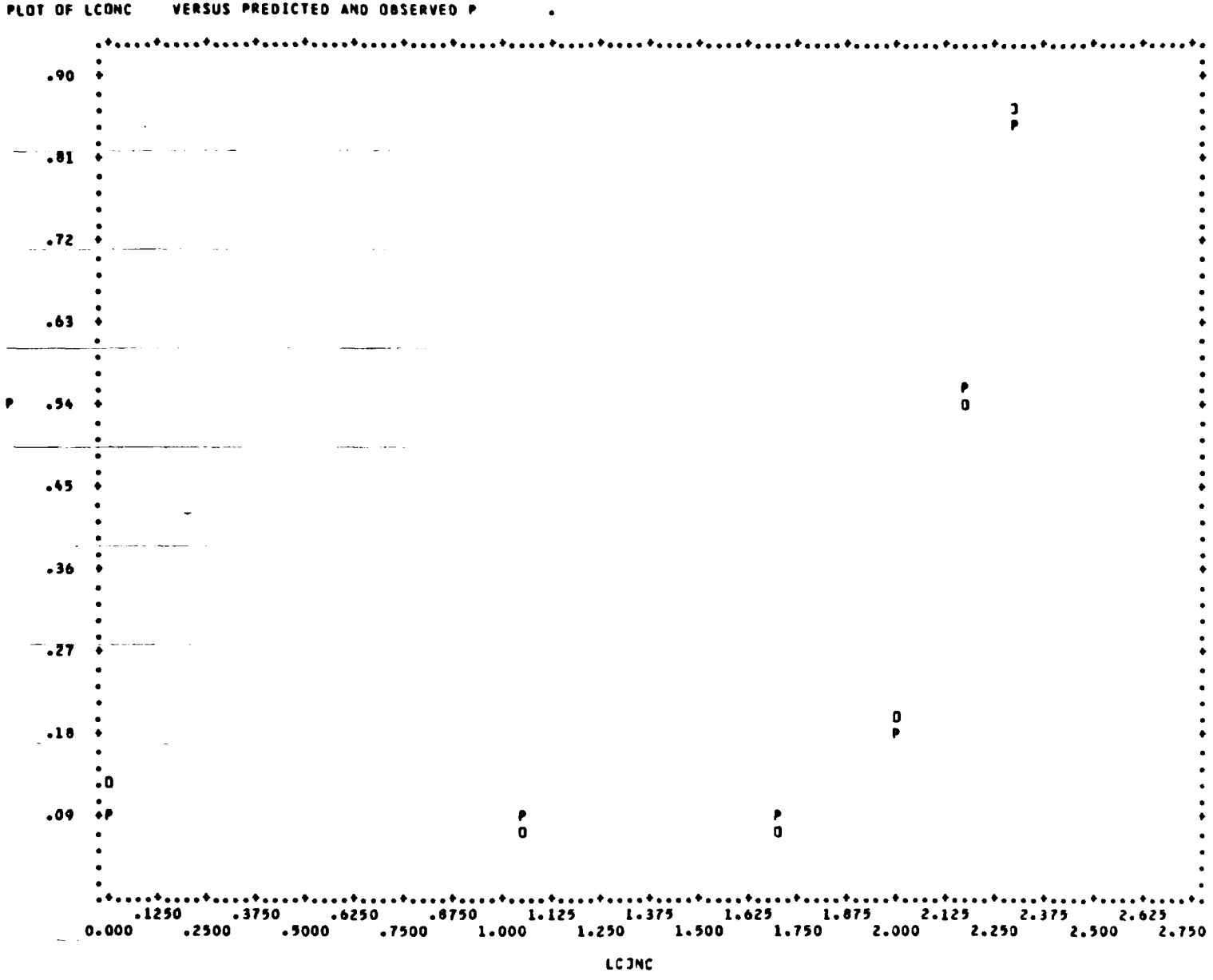


Figure X.13. Observed and predicted values from fit in Figure X.11 versus logarithmic concentration ($\log_{10}$ (1 + conc)).

XI. TESTING FOR CONCENTRATION RELATED EFFECTS ON REPRODUCTION AND LENGTH

A. INTRODUCTION

In Sections IX and X we considered the comparison of mortality rates across groups. Section IX was concerned with hypothesis testing and multiple comparison procedures while Section X was concerned with fitting dose response curves and estimating safe concentrations based on the percentiles of these curves. A directly analogous situation holds for quantitative responses such as length and weight. In this section we consider hypothesis testing and multiple comparison procedures and in the next section we consider dose response estimation and associated inferences based on multiple regression models.

While the procedures employed to analyze the length and reproduction responses are for the most part similar to those used to analyze mortality, there are a number of conceptual and technical differences in the problems. Several of these are discussed below.

An important issue is related to mortalities censoring the nonlethal responses. That is, measurements such length, reproduction, weight, brood size, etc. obviously can only be determined on survivors. Thus the daphnids upon which the determinations of nonlethal doses are based are not chosen at random, but rather are the hardiest in the groups, since they survived. This can potentially give rise to biased comparisons among groups since at high concentrations the weaker daphnids would be killed off while at low concentrations these weaker daphnids would survive and register inferior lengths or reproduction. This could mask dose response effects or even show reverse effects (e.g., greater average lengths among the survivors at the high concentrations than at the low).

There are no completely satisfactory ways to eliminate such potential biases. One partial solution is to regard the responses in a heirarchy. Mortality would be a first order effect. Any concentration which results in "substantial" mortality would be considered unsafe, irrespective of any other responses. Sublethal effects such as reductions in length or reproduction would be considered second order. Among concentrations which pass the mortality screen, any that result in biologically and statistically significant nonlethal effects would be considered unsafe. Thus before testing hypotheses about, or fitting dose response curves to length or reproduction responses, we delete those groups with "substantial" mortality. We have not precisely defined "substantial" for this purpose. An operational definition such as 20, 30, 50, etc., percent increases in mortality above background might be used. Gelber, et al [28], suggest deleting all groups at concentrations beyond the MATC "...to achieve comparable numbers of survivors per tank...". Rather then adopting such a formal approach, we made individual decisions on a test-by-test basis. In particular, we deleted the length and reproduction results from Group 7 in LeBlanc's Test A, from Groups 5, 6, 7, 8 in Adams' Selenium test, and from

Group 6 in Goulden's Isophorone test. The consequences on interpretation of results of deleting these high mortality groups from comparisons of nonlethal effects requires further discussion and study among toxicologists.

Another difference between the inference problems associated with mortality and those associated with nonlethal responses is related to the monotonicity or lack of monotonicity of the response levels with increasing concentration. The inference procedures for mortality are based on the assumption that mortality rate increases (or at least does not decrease) with increasing concentration. The measure of association test, the Cochran-Armitage Test, Williams' Test, and the probit model all require monotonicity. Such monotonicity does not necessarily hold for length and reproduction. Figures II.11-II.14 and II.21-II.24 show that reproduction levels in LeBlanc's Tests A and B and in Goulden's Isophorone test first increase and then decrease as concentration is increased. Figure II.29 shows that average lengths in LeBlanc's Test B first increase and then decrease as concentration is increased. Thus the inference procedures used must be valid for nonmonotone trends.

Inference procedures for mortality are generally based on the binomial distribution. Such procedures tacitly assume that the variances of responses are certain specified functions of the means. Thus variance estimates need not be supplied. However for quantitative responses such as length and reproduction, the comparable procedures are based on regression analysis and analysis of variance which do require estimates of variability. The question thus arises as to how these variability estimates will be calculated. They should simultaneously account for possible beaker-to-beaker heterogeneity within groups, yet utilize all the information in the data. In Chapman's and in Goulden's data sets, length and reproduction responses are measured on daphnids housed one per beaker. Thus variability of responses is estimated based on observed individual daphnid-to-daphnid variability per group. In LeBlanc's and in Adams' data sets, reproduction responses are measured on a per beaker basis. Thus the basic responses are numbers of offspring per beaker, normalized to reflect the numbers of surviving adults. Thus variability is estimated based on observed beaker-to-beaker variability per group. The situation for the length responses in LeBlanc's tests is a bit more complex. Daphnids are multiply housed within beakers; however lengths are measured on individual daphnids. Thus we would like to somehow use the variability among individual responses. Yet these individual responses are possibly correlated due to beaker-to-beaker heterogeneity within groups. A scheme for pooling estimates of variability among beaker averages within groups with estimates of variability among daphnids within beakers was discussed in Subsection V-D. This approach uses variance estimates based on variability among beaker averages but augments the degrees of freedom to reflect the information about daphnid-to-daphnid variability within groups. The pooling of information has the greatest impact when there are few degrees of freedom for variance estimation based on beaker averages within groups and there is little beaker-to-beaker heterogeneity.

We now consider various hypothesis testing and multiple comparison procedures to compare length and reproduction responses across groups. Each of the hypothesis testing procedures considered in Section IX for mortality responses has directly analogous counterparts appropriate for

quantitative responses such as length and reproduction. Chi square tests correspond to analysis of variance tests. Measure of association tests correspond to inferences about the correlation coefficient. The Cochran-Armitage Test corresponds to a test based on straight-line regression trend. Dunnett's and Williams' multiple comparison procedures carry over directly. However, since the trends in responses are not necessarily monotone we do not carry out the test procedures in the same sequential manner that we did for mortality responses--namely peeling off the highest treatment group and retesting after each significant result. Also, Dunnett's procedure is used for multiple comparisons rather than Williams' procedure.

B. REPRODUCTION RESPONSES--ANALYSIS OF VARIANCE AND MULTIPLE COMPARISON PROCEDURES

The analysis of variance procedures are based on the one-way analysis of variance model. We used this procedure in Subsection VI-C when we were looking for outliers. See for example Figures VI.3, VI.6, VI.9, VI.12, VI.15. The tests we use in this section are directly analogous except that we delete some groups or some individual responses because of excess mortality or because of outliers. The beaker is the basic response unit. We present several illustrative examples below.

LeBlanc Test A--Water Control Group--Group 7 Deleted--Analysis of Variance

There are 4 treatment groups and a control group, with 4 beakers per group. There are thus 4 degrees of freedom for comparisons among groups and 15 degrees of freedom with which to estimate error. The analysis of variance test is a "shotgun test", in the same manner as the chi square test for mortality responses. The analysis of variance table is

21 Day Cumulative Offspring Per Surviving Daphnid

Source	D.F.	Sum of Squares	Mean Square	F Ratio
Between Groups	4	4787.3	1196.825	1.969
Within Groups	15	9115.25	607.68	
Adjusted Total	19	13902.55		

The upper 90 percent point of the F distribution with degrees of freedom 4 and 15 is 2.27. Thus there is no statistical evidence with this test of differences in reproduction rates among groups, after excluding the very high mortality group.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted--Analysis of Variance

The framework is the same as that above, except that the solvent control group responses are substituted for the water control group responses. The analysis of variance table is

21 Day Cumulative Offspring Per Surviving Daphnid

Source	D.F.	Sum of Squares	Mean Square	F Ratio
Between Groups	4	6002.7	1500.675	2.633
Within Groups	15	8550.25	570.017	
Adjusted Total	19	14552.95		

The upper 90 and 95 percent points of the F distribution with degrees of freedom 4 and 15 ar 2.36 and 3.06 respectively. Since the observed F ratio falls between these values we conclude that there is a suggestion, but not strong statistical evidence of average differences in reproduction rates among groups.

Note that there is a somewhat different outcome with this test, depending on whether the solvent or the water control group is used.

LeBlanc Test B--Combined Control Groups--Analysis of Variance

The framework is similar to that for Test A except that no treatment groups are deleted and the solvent and water control group responses are combined into a single group. The analysis of variance table is

21 Day CUmulative Offspring Per Surviving Daphnid

Source	D.F.	Sum of Squares	Mean Square	F Ratio
Between Groups	5	2542.714	508.54	1.149
Within Groups	22	9738.000	442.64	
Adjusted Total	27	12280.714		

The upper 90 percent point of the F distribution with degrees of freedom 5 and 22 is 2.13. Thus there is no statistical evidence based on this test of differences in average reproduction rates among groups.

Goulden Isophorone Test--Group 6 Deleted--Outlier in Group 1 Deleted--Analysis of Variance

These comparisons are based on the reproduction responses from the individually housed daphnids in each group that survived to the end of the test. There were 7 such daphnids per group at the outset of the test. All but one of these daphnids survived. Beaker 5 in Group 1 was determined in Subsection VI-C to have an outlying response and that was also deleted. The comparison is thus based on the individual responses from 33 daphnids. The analysis of variance table is

21 Day Cumulative Offspring for Each Daphnid Surviving to the End of the Test

Source	D.F.	Sum of Squares	Mean Square	F Ratio
Between Groups	4	26414.46	6603.62	35.32
Within Groups	28	5234.45	186.95	
Adjusted Total	32	31648.91		

This F ratio is of course highly statistically significant. There is thus strong statistical evidence of average differences in reproduction levels among groups. This is not very surprising, based on the appearance of Figure II.22.

The previous discussion in this subsection has been based on one-way analysis of variance tests. These tests are overall, "shotgun" tests and so are not the most sensitive to the types of departures from homogeneity that are of most importance in toxicity tests. In particular, it is often desired to carry out a series of pairwise treatment group-control group comparisons and determine which treatment group responses are significantly different then the control group response. Since such pairwise comparisons focus in on the effects of interest, they are more sensitive tests then the overall chi square test.

Two commonly used procedures for carrying out treatment-group control multiple comparisons are Williams' test and Dunnett's test. See Williams [17,18] and Dunnett [15,16] for detailed descriptions of their use and for appropriate tables. We have applied Williams' test to study the mortality data. Williams' test assumes that the response curve varies monotically with concentration. Since this assumption is not necessarily valid for reproduction, we use Dunnett's procedure instead. We present several examples below, based on the same data sets as those discussed above with the analysis of variance tests.

LeBlanc Test A--Water Control Group--Group 7 Deleted--Dunnett's Test

The numbers of beakers and the average cumulative offspring per beaker within each group are

Group	1	3	4	5	6
N	4	4	4	4	4
Average	102.5	128.25	129	139.75	101.25

The standard errors of these averages, based on the analysis of variance fit are $(\hat{\sigma}^2/4)^{1/2}=(607.68/4)^{1/2}=12.33$. Let $\bar{X}_i$ denote the average in the i-th group.

We apply Dunnett's procedure to determine which groups have (statistically) significantly lower average reproduction then the control group. We declare the group i average reproduction to be significantly lower than the control average if

$$\bar{X}_i-\bar{X}_1<-\bar{t}(2\hat{\sigma}^2/4)^{1/2}$$

The factor $\bar{t}$ is obtained from Dunnett's tables of one-sided factors and is derived under the assumption of equal group sample sizes. In this example, $\hat{\sigma}^2$ is estimated with 15 degrees of freedom. The value of t corresponding

to 15 degrees of freedom, 4 treatment groups, and α=0.05 is $\bar{t}$=2.36. Therefore the critical value is

$$\bar{X}_1-\bar{t}(2\hat{\sigma}^2/4)^{1/2}=102.5-(2.36)(17.431)=61.36$$

Thus none of the treatment group averages are significantly lower than the control group average. Thus the results of this test agree with those based on the analysis of variance test.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted--Dunnett's Test

The numbers of beakers and the average cumulative offspring per beaker within each group are the same as shown above, except that Group 2 is substituted for Group 1. The values for Group 2 are N=4 and Average=154. The standard errors of these averages, based on the analysis of variance fit are $(\hat{\sigma}^2/4)^{1/2}=(570.017/4)=11.94$. Let $\bar{X}_i$ denote the average in the i-th group. The group i average reproduction is significantly lower than the control average if

$$\bar{X}_i-\bar{X}_2<-\bar{t}(2\hat{\sigma}^2/4)^{1/2}$$

As above, the estimated variance $\hat{\sigma}^2$ has 15 degrees of freedom. The $\bar{t}$ factor, corresponding to 15 degrees of freedom, 4 treatment groups, and α=0.05 is $\bar{t}$=2.36. Thus the critical value is

$$\bar{X}_2-\bar{t}(2\hat{\sigma}^2/4)^{1/2}=154-(2.36)(16.88)=114.16.$$

This implies that the average reproduction rate in Group 6 is significantly smaller then that in the solvent control group at α=0.05. We can repeat this test after deleting Group 6 and adjusting $\bar{t}$ to correspond to 3 treatment groups. The new critical value is 154-(2.24)(16.88)=116.19. Thus no other groups have significantly lower average reproduction then the solvent control group.

We come to different conclusions, depending on whether the water control group or the solvent control group is used for comparison. This raises important conceptual problems, as discussed in Section VII.

LeBlanc Test B--Combined Control Groups--Dunnett's Test

There are 5 treatment groups and a (combined) control group. The numbers of beakers and the average cumulative offspring beaker are

Group	1 and 2	3	4	5	6	7
N	8	4	4	4	4	4
Average	109.25	116	134.5	128.25	119.25	108

The estimated error variance, based on the analysis of variance fit, is $\hat{\sigma}^2=442.64$ with 22 degrees of freedom. Let $\bar{X}_i$ denote the average in the i-th group. The group i average reproduction is significantly lower than the control average if

$$\bar{X}_i-\bar{X}_o<-\bar{t}\ \hat{\sigma}(1/N_i+1/N_o)^{1/2}$$

where $\bar{X}_o, N_o$ are the average reproduction and the number of beakers in the combined control group. In this example, $N_i=4, N_o=8$.

The tabulations of Dunnett's factors are based on the assumption of equal sample sizes in each group. Since the control group has twice as many beakers as any of the treatment groups, the appropriate values of t must be obtained from tables of the multivariate t distribution. Krishnaiah [29], pp 789-800 has published tables of this distribution. For $i\neq j$ the correlation between X_i-X_o and X_j-X_o is 1/3. We must enter the tables of the multivariate t distribution at 22 degrees of freedom, 5 groups, and correlation $\rho=0.33$ to find the upper $\alpha=0.05$ point. Interpolating in Krishnaiah's tables between $\rho=0.2$ and $\rho=0.4$ yields $\bar{t}=2.43$.

Thus the critical value is

$$\bar{X}_o-\bar{t}\hat{\sigma}(1/N_i+1/N_o)^{1/2}=109.25-(2.43)(21.04)(0.61)=77.94$$

Since the control group has one of the lowest reproduction rates among all the groups, there are obviously no groups significantly lower then the control.

Goulden Isophorone Test--Group 6 Deleted--Outlier in Group 1 Deleted--Dunnett's Test

The numbers of beakers and the cumulative offspring per beaker within each group are

Group	1	2	3	4	5
N	6	6	7	7	7
Average	67.667	86.833	90.857	69.000	13.286

Each beaker contains just one daphnid and attention is confined to those daphnids that survived to the end of the test. The estimated error variance, based on the analysis of variance fit, is $\hat{\sigma}^2=186.95$ with 28 degrees of freedom. Let $\bar{X}_i$ denote the average in the i-th group. The group i average reproduction is significantly lower than the control if

$$\bar{X}_i-\bar{X}_1<-\bar{t}\hat{\sigma}(1/N_i+1/N_1)^{1/2}$$

Even though the sample sizes are not exactly equal, they are almost equal, and so we use Dunnett's factors as an approximation. Thus the critical value is

$$\begin{aligned}\bar{X}_1-\bar{t}\hat{\sigma}(1/N_i+1/N_1)^{1/2} &= 67.667-(2.26)(13.67)(1/N_i+1/6)^{1/2}\\ &= 49.83 \text{ if } i=2\\ &= 50.48 \text{ if } i=3,4,5\end{aligned}$$

Thus Group 5 has a significantly smaller reproduction rate then the control group. Obviously none of the other groups do.

C. LENGTH RESPONSES--ANALYSIS OF VARIANCE AND MULTIPLE COMPARISON PROCEDURES

The situation for comparisons of lengths across groups is similar to that for comparisons of reproduction, except that lengths are measured on a per daphnid basis whereas reproduction is measured on a per beaker basis. Thus unless there is just one daphnid per beaker, less information is obtained on reproduction than on lengths.

A basic technical difficulty associated with directly analyzing the individual length determinations is due to the correlations among responses from daphnids in the same beaker. This intercorrelation results from the beaker-to-beaker heterogeneity within groups. A common practice is to summarize the individual length measurements within each beaker by the average length and then analyze the averages on a per beaker basis, as was previously done for the reproduction responses. This is essentially equivalent to fitting a two-way nested analysis of variance model to the data, as was explained and illustrated in Subsection IV-B, and using the mean square for variation among beakers within groups as the error term for making inferences about treatment effects.

If the mean square for beakers within groups is no greater than the mean square for variation among daphnids within beakers then the two mean squares are sometimes pooled and used as a common error term. The latter

mean square usually has many more degrees of freedom than the former. In Subsection V-D we discussed a scheme for pooling information from these two mean squares in a continuous manner. Basically, the mean square for beakers within groups is used as the error yardstick but information from the within beaker mean square is used to augment the degrees of freedom in a continuous manner. The closer are the two mean squares, the greater are the degrees of freedom. See Subsection V-D for details. The results of this pooling procedure applied to the length responses from LeBlanc's Tests A and B are:

Test A: $\hat{\sigma}^2$=0.2722 with 98 degrees of freedom (excluding Group 7)
Test B: $\hat{\sigma}^2$=0.2707 with 141 degrees of freedom

We use these error estimates in subsequent analyses.

The analysis of variance procedures are based on the two-way nested analysis of variance model. Although we should separate the control groups in Test A, combine the control groups in Test B, and re-estimate the error variances from the modified data, we utilize the analysis of variance fits shown in Subsection IV-B to illustrate the calculation of the analysis of variance tests. Although these analyses of variance do not test quite the right hypotheses, they do illustrate the appropriate methodology.

LeBlanc Test A--Analysis of Variance Test

From Subsection IV-B, mean square between groups =1.1163 with 5 degress of freedom. Mean square for beakers within groups =0.2722 with 98 degrees of freedom. Thus

$$F = \frac{1.1163}{0.2722} = 4.101 \text{ with 5 and 98 degrees of freedom.}$$

This F ratio is statistically significant at α=0.002. Thus there is strong statistical evidence of differences in average lengths among groups. From Figure II.26 it appears that average length in the solvent control group is somewhat greater than the average lengths in the other groups.

LeBlanc Test B--Analysis of Variance Test

From Subsection IV-B, mean square between groups =0.2915 with 6 degrees of freedom. Mean square for beakers within groups =0.2707 with 141 degrees of freedom. Thus

$$F = \frac{0.2915}{0.2707} = 1.077 \text{ with 6 and 141 degrees of freedom.}$$

This F ratio is significant at α=0.38. Thus there is no statistical evidence of differences in average lengths among groups. The relatively small lengths in Group 7 do not show up in this test. See Figure II.29 for a graphical display of the group-to-group variation in lengths.

We now carry out pairwise comparisons of treatment group and control group average responses using Dunnett's test. Lengths were measured by LeBlanc in Tests A and B and by Chapman in his Beryllium test. For the LeBlanc data we use the variance estimates and degrees of freedom arrived at in Subsection V-D, namely $\hat{\sigma}^2$=0.2722 with 98 degrees of freedom for Test A (excluding Group 7) and $\hat{\sigma}^2$=0.2707 with 141 degrees of freedom for Test B.

Before considering the examples below, we must address a technical issue concerning the estimation of the average responses within each group. There are two components of variation, a beaker-to-beaker component and a daphnid-to-daphnid component within beakers. The question arises as to whether we should calculate a simple average of all the daphnid responses within each group, calculate the average of the average beaker responses, or perhaps some compromise between these two averages. In the balanced situation where each beaker has the same number of daphnids these two averaging processes yield the same results and there is no ambiguity. However in the unbalanced case those averaging processes can yield very different results. In LeBlanc's Tests A and B, most of the beakers within groups have between 15 and 20 daphnids and so averaging processes based on the balanced case should be quite reasonable. We thus use unweighted averages of the individual responses in these groups. However the situations in Group 6 of Test A and in Group 7 of Test B are different. The sample sizes in the four beakers in Group 6 of Test A are 18, 16, 5, and 1 and sample sizes in the four beakers in Group 7 of Test B are 3, 14, 12, and 1. These are very highly imbalanced. Thus how should the average responses be estimated, in these groups? A complete answer to this question is a research problem in its own right and we will not attempt that here. However an intuitively reasonable approach would be to calculate that average of the responses that most precisely estimates mean length within the group. Let σ_β^2, σ_e^2 denote the variance components due to beakers and daphnids respectively and let $\rho=\sigma_\beta^2/\sigma_e^2$. Suppose there are J beakers, N_j daphnids within the j-th beaker, and the average response in the j-th beaker is $\bar{X}_j$. Then $\mathrm{Var}(\bar{X}_j)=\sigma_\beta^2+\sigma_e^2/N_j$. We consider estimates of the form

$$\hat{\mu}=\Sigma_{j=1}^{J} w_j\bar{X}_j \quad \text{with } w_j\geq 0,\ \Sigma_{j=1}^{J} w_j=1$$

We choose the weights so as to minimize the variance of $\hat{\mu}$. This is a Lagrange multiplier problem. The solution is to choose w_j proportional to $N_j/(1+N_j\rho)$. As ρ approaches o, we tend to average individual responses and as ρ approaches infinity we tend to average group averages. The general situation is a compromise between these two extremes. In the LeBlanc data sets, ρ is estimated to be 0.09 in Test A and 0.06 in Test B. Since these values of ρ are small, we use the simple averages of the individual observations for weighting purposes, although the "optimum" weights call for assigning lower weights to the beakers with relatively large numbers of

daphnids. The variance of this simple average is $\sigma_e^2/\Sigma N_j+\sigma_b^2\Sigma N_j^2/(\Sigma N_j)^2$. We use this variance expression for the two groups with highly unbalanced samples sizes. For the other groups, we use variance expressions appropriate for the balanced case as these should be reasonable approximations. A more rigorous analysis could be based on maximum likelihood methods, but even here the theory holds only asymptotically.

We now consider several examples.

LeBlanc Test A--Water Control Group--Group 7 Deleted--Dunnett's Test

The numbers and average lengths of surviving daphnids within each group and the standard errors of the mean are

Group	1	3	4	5	6
N_j	68	70	73	73	40
Average	4.9676	4.9743	5.1027	4.9658	5.0000
Std Error	0.063	0.062	0.061	0.061	0.080

The standard error estimates in Group 1-5 are based on the square root of the beakers within groups mean square shown in Subsection V-D, divided by N_j. The standard error estimate in Group 6 is based on the expression in the paragraph preceding the example. We associate 98 degrees of freedom with these standard error estimates. Let $\bar{X}_i$ denote the average in the i-th group. The group i average length is significantly lower than the control average if

$$\bar{X}_i-\bar{X}_1<-\bar{t}(\text{std err}_i^2+\text{std err}_1^2)^{1/2}$$

The $\bar{t}$ factor, corresponding to 98 degrees of freedom, 4 treatment groups, and α=0.05 is $\bar{t}$=2.19. Thus the critical values are:

Group 3 vs Control: $4.9676-(2.19)(0.063^2+0.062^2)^{1/2}=4.774$
Group 4 vs Control: $4.9676-(2.19)(0.063^2+0.061^2)^{1/2}=4.776$
Group 5 vs Control: $4.9676-(2.19)(0.063^2+0.061^2)^{1/2}=4.776$
Group 6 vs Control: $4.9676-(2.19)(0.063^2+0.080^2)^{1/2}=4.745$

Since none of the group averages are below their critical values, we conclude that there is no statistical evidence that any of the treatment groups have significantly lower average lengths then the water control group.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted--Dunnett's Test

The numbers and average lengths of surviving daphnids within each group and the standard errors are the same as those shown above, except that Group 2 is substituted for Group 1. The values for Group 2 are ΣN_j=77, Average=5.2753, Std Error=0.059. The group i average length, $\bar{X}_i$, is significantly lower then the solvent control group average if

$$\bar{X}_i - \bar{X}_2 < -\bar{t}(\text{std err}_i^2 + \text{std err}_2^2)^{1/2}$$

The $\bar{t}$ factor is again 2.19. Thus the critical values are:

Group 3 vs Control: $5.2753-(2.19)(0.059^2+0.062^2)^{1/2}=5.088$
Group 4 vs Control: $5.2753-(2.19)(0.059^2+0.061^2)^{1/2}=5.089$
Group 5 vs Control: $5.2753-(2.19)(0.059^2+0.061^2)^{1/2}=5.089$
Group 6 vs Control: $5.2753-(2.19)(0.059^2+0.080^2)^{1/2}=5.058$

The averages in Groups 3, 5, and 6 are below their critical values and so these averages are significantly lower then the control group average.

Since Groups 3, 5, and 6 differ significantly from the control group, we recalculate the critical value for Group 4 based on just one treatment group (i.e., the usual Student's t distribution). The t factor changes from 2.19 to 1.66. Thus the critical value for Group 4 vs Control is

$$5.2753-(1.66)(0.059^2+0.061^2)^{1/2}=5.134$$

The Group 4 average is lower then this value and thus is also significantly lower then the control.

We conclude that all of the treatment groups have significantly lower average lengths than the solvent control group, at the α=0.05 level of significance.

Note that we have arrived at diametrically opposite conclusions, depending on whether the water control group or the solvent control group is used for comparison. This situation is clearly seen in Figure II.26. As remarked previously, these contradictory conclusions lead to important conceptual problems in interpreting the results of the test.

LeBlanc Test B--Combined Control Groups--Dunnett's Test

The numbers and average lengths of surviving daphnids in each group and the standard errors of the means are

Group	1 and 2	3	4	5	6	7
N_j	139	76	75	67	72	30
Average	4.8907	4.8645	4.9360	4.8358	4.8333	4.7267
Std Err	0.044	0.060	0.060	0.064	0.061	0.088

The standard error estimates in the control group and in treatment Groups 3-6 are based on the square root of the beakers within groups mean square shown in Subsection V-D, divided by ΣN_j. The standard error estimate in Group 7 is based on the expression in the paragraph preceding the first example in this series. We associate 141 degrees of freedom with these standard error estimates. Let $\bar{X}_i, \bar{X}_o$ denote the average in the i-th treatment group and in the combined control group respectively. The group i average length is significantly lower than the control average if

$$\bar{X}_i - \bar{X}_o < -\bar{t}(\text{std err}_i^2 + \text{std err}_o^2)^{1/2}$$

The $\bar{t}$ factor, corresponding to 141 degrees of freedom, 5 treatment groups, and $\alpha=0.05$ is obtained from Krishnaiah's tables of the multivariate t distribution, pp 789-800. Since the number of beakers and daphnids in the combined control group is about twice that in the treatment groups (except for Group 7), the correlation between $\bar{X}_i - \bar{X}_o$ and $X_j - X_o$ is about 1/3 for $i \neq j \neq 7$. We enter Krishnaiah's tables at 141 degrees of freedom, 5 groups, and correlation $\rho=0.33$ to find the upper $\alpha=0.05$ point. Interpolating between $\rho=0.2$ and $\rho=0.4$ yields $\bar{t}=2.37$. (The table actually goes up to only 35 degrees of freedom.) Thus the critical values are:

Group 3 vs Control: $4.8907-(2.37)(0.044^2+0.060^2)^{1/2}=4.714$
Group 4 vs Control: $4.8907-(2.37)(0.044^2+0.060^2)^{1/2}=4.714$
Group 5 vs Control: $4.8907-(2.37)(0.044^2+0.064^2)^{1/2}=4.707$
Group 6 vs Control: $4.8907-(2.37)(0.044^2+0.061^2)^{1/2}=4.712$
Group 7 vs Control: $4.8907-(2.37)(0.044^2+0.088^2)^{1/2}=4.658$

Since none of the group averages are less than their critical values, we conclude that there is no statistical evidence that any of the treatment groups have significantly lower average lengths than the combined control group. The appearance of Figure II.29 bears out this conclusion, except for Group 7 which is a bit lower. However it is not significantly lower. The analysis of variance test gave the same conclusions.

XII. DOSE RESPONSE CURVE ESTIMATION-MULTIPLE REGRESSION ANALYSIS OF LENGTH AND REPRODUCTION DATA

A. INTRODUCTION

In the previous section we considered hypothesis testing and multiple comparison procedures to test for the presence of differences among groups in average length and reproduction. These testing procedures are analogous to those considered in Section IX for mortality responses. In this section we consider fitting multiple regression dose response models to the length and reproduction data to estimate the concentrations, C_L, which result in reductions of L relative to the control group levels. These multiple regression procedures for quantitative responses are directly analogous to the probit analysis dose response model for qualitative responses that was considered in Section X.

The conceptual distinctions between the hypothesis testing procedures of the previous section and the multiple regression procedures of this section are the same as the distinctions between corresponding procedures for mortality data that are discussed at the beginning of Section X. In brief, inferences based on the multiple regression procedures incorporate biological significance as well as statistical significance and tend to result in tighter confidence bounds on safe concentrations (and thus in more liberal lower bounds) as the amount and precision of the data increase. We feel that this approach to inference has more appeal than hypothesis testing.

The two principal technical differences between the multiple regression models appropriate for studying length and reproduction responses and the probit model appropriate for studying mortality responses relate to the nonmonotone nature of the trends in length and reproduction and the need to supply estimates of variability. Both of these considerations were discussed in Subsection XI-A and so need not be repeated here.

Feder and Collins [1] discuss fitting multiple regression models to weight gain data from early life stage tests with fathead minnows, in Subsection XVII-C of their report.

We now discuss the specific models that were fitted to the data. Let $x \equiv \log_{10}$ (concentration). (Since the regression models will be fitted only to the treatment groups, there is no problem with the logarithm of 0.) Let m,s denote location and scale standardization factors respectively. (Note that these are <u>not</u> necessarily the mean and standard deviation.) Let $v \equiv (x-m)/s$ denote the standardized version of x and let y denote the response (average length or cumulative reproduction per surviving adult). Let I denote an indicator function of the treatment groups. That is, I=1 for treatment groups and I=0 for control groups. The models fitted are:

Straight Line

$$Y=\mu+\beta_0 I+\beta_1 Iv+\varepsilon$$

Quadratic

$$Y=\mu+\beta_0 I+\beta_1 Iv+\beta_2 Iv^2+\varepsilon$$

In these model specifications μ represents the control group average, β_0, β_1, and (possibly) β_2 represent the coefficients of the polynominal trend in the region of the treatment groups. By the parameterization of the model, the polynominals $\beta_0+\beta_1 v$ or $\beta_0+\beta_1 v+\beta_2 v^2$ represent the difference between the treatment group expected response at v and the control group average. Since the average length and reproduction levels in the control groups generally differ from zero, they must be adjusted for in our models. We wish to estimate the value of v where

$$\beta_0+\beta_1 v=c$$

or

$$\beta_0+\beta_1 v+\beta_2 v^2=c$$

and place a confidence interval on this value. The theory underlying the point and confidence interval estimation procedure is discussed in Appendix AXII.1. A computer program to carry out the calculations is described in Appendix AXII.2.

We now discuss fitting regression models to the reproduction responses and to the length responses in turn.

B. REPRODUCTION RESPONSES

In this subsection we fit multiple regression models to study trends in cumulative reproduction as concentration increases. We fit linear or quadratic models, as appropriate, to the treatment group data and estimate concentrations resulting in specified decreases from the control group in average reproduction. The same approach could of course be extended to fit cubic models, exponential models, etc., to the data. Such extensions have not been explored, but are straightforward. Since reproduction responses are measured on a per beaker basis, there are no questions of or complications due to beaker-to-beaker heterogeneity. In the Chapman and Goulden tests, reproduction was measured on individually housed daphnids. Only the daphnids that survived to the end of the test were included in the analysis. We now consider the details of the fits to the various data sets.

LeBlanc Test A--Water Control Group--Group 7 Deleted

There are four treatment groups and a control group. The concentrations corresponding to the lowest two treatment groups, Groups 3 and 4, are virtually identical. We are thus fitting a quadratic model to essentially three distinct concentrations (recall that we have excluded the control group). The model fitted is

$$Y=\mu+\beta_0 I+\beta_1 Iv+\beta_2 Iv^2+\varepsilon$$

Where $v=(x+2.257)/0.311$ (i.e.,m=-2.257 and s=0.311). The results of this fit are shown in Figure XII.1. The observed and predicted values are plotted in Figure XII.2 vs $\log_{10}$ (concentration). The estimated regression coefficients and their estimated variance covariance matrix are shown in Figure XII.1. Figure XII.2 shows a nonlinear trend in average response within the range of the treatment groups. The quadratic coefficient in Figure XII.1 is marginally significant (α=0.08). We thus base estimates of acceptable concentratio s on the quadratic model.

The output from CONFINT applied to this model is shown in Figures XII.3,4. We calculate point and confidence interval estimates of concentrations corresponding to 10 percent and 20 percent reductions in average reproduction, relative to the control group. Since the estimated cumulative reproduction rate in the water control group is 102.5 offspring per adult, the values of c specified are -10.25 and -20.5 respectively. Two roots are shown in each Figure. The smaller roots, -2.76 and -2.81 in logarithmic concentration units, are obviously inappropriate based on the appearance of Figure XII.2. We thus confine attention to the larger roots. These correspond to roots no. 1 in the two outputs. From the bottoms of Figures XII.3,4 we read these as

10 percent decrease: $\hat{conc}$=0.0168 95 percent conf interval (0.0107, 0.0262)
20 percent decrease: $\hat{conc}$=0.0187 95 percent conf interval (0.0116, 0.0300)

These point estimates both exceed the concentration in the highest treatment group. Thus they represent extrapolations beyond the range of the data, with the consequent possibility of extrapolation biases. However, since the amount of extrapolation is very little, this should not be a problem in this example.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted

The framework is the same as in the previous example except that the solvent control group (group 2) is substituted for the water control group. There are again essentially three distinct treatment group concentrations. The model fitted is the same as that in the previous example. The

standardization factors are also the same. The results of this fit are shown in Figures XII.5 to XII.8. The quadratic fit to the treatment group responses is the same as in the previous example but the control group average is somewhat greater (154.0 vs 102.5).

The output from CONFINT applied to this model is shown in Figures XII.7,8. Since the estimated cumulative reproduction rate in the solvent control group is 154.0 offspring per adult, the values of c corresponding to 10 percent and 20 percent reductions are -15.4 and -30.8 respectively. As in the previous example, the larger of the two roots in each output are the ones of biological importance. These correspond to roots no. 1 in the outputs. From the bottoms of Figures XII.7,8 we read these as

10 percent decrease: $\hat{conc}$=0.0077 95 percent conf interval (0.0023, 0.0254)
20 percent decrease: $\hat{conc}$=0.0110 95 percent conf interval (0.0064, 0.0189)

Note that the confidence interval for the concentration associated with the 10 percent decrease is much wider than that associated with the 20 percent decrease. This is because the point estimate, x=-2.11, lies near the stationary point of the curve. Thus the slope of the curve is very gentle in this region and so small changes in the curve correspond to large changes in concentration.

Comparing the concentrations associated with 10 percent and 20 percent decreases based on the solvent control group with those based on the water control group, we see substantially lower point estimates with the solvent control group than with the water control group. The confidence interval associated with 10 percent decreases from the solvent control group response is so wide as to be useless. The other three confidence intervals span a factor of three in concentration levels and so are also too wide to provide very precise estimates. However, the confidence interval associated with 20 percent decreases from the solvent control group is substantially lower than that associated with 20 percent decreases from the water control group. (The endpoints for the water control interval are 60 percent to 80 percent higher than those for the solvent control interval.)

It appears that although the confidence intervals are rather wide, they are precise enough to conclude that qualitatively different results are obtained depending on whether the water control group or the solvent control group is used for comparison. This results in important interpretational ambiguity.

Chapman-Beryllium Test

The data consist of water and solvent control groups and six treatment groups. Since the daphnids were individually housed, individual reproduction determinations were made and these are used for analysis. Responses only from daphnids that survived to the end of the test are included in the

analysis. The results from the two control groups were combined for comparison with the treatment groups. The results of the analysis are shown in Figures XII.9 to XII.13.

The quadratic model that was discussed previously was fitted to the data. The independent variable, v, was defined as $v=(x-1.276)/0.551$ (i.e., $m=1.276$ and $s=0.551$). The results of the fit are shown in Figure XII.9. The observed and predicted values are plotted in Figure XII.10 vs $\log_{10}$ (concentration). The quadratic term is nonsignificant ($\hat{\beta}_2=-6.042$ with estimated standard error=8.433). There appears to be no trend in average reproduction with increasing concentration displayed in Figure XII.10. The quadratic term was deleted from the model and the straight line model $Y=\mu+\beta_0 I+\beta_1 Ix+\varepsilon$ was fitted to the data. The results are shown in Figure XII.11. The linear term is nonsignificant ($\hat{\beta}_1=5.816$ with estimated standard error 13.165). We thus conclude that there is no significant trend in reproduction within the range of the treatment group concentrations. It is therefore meaningless to estimate concentrations associated with specified reductions from the control response. If we go ahead anyway and formally carry out the inference, we get results such as shown in Figure XII.12. The confidence interval on the concentration associated with a 10 percent reduction from the control group average ranges from 0 to infinity. It is thus of course meaningless.

Because of the lack of trend among the treatment group responses, observed in the regression outputs and in the display in Figure XII.10, we calculate an overall level of reproduction within the treatment groups for comparison with the reproduction rate in the control groups. In particular for the N=45 surviving daphnids in treatment groups 3-8, the mean and standard deviation are

$$\bar{Y}_1=92.844 \qquad \hat{\sigma}_1=43.242$$

For the N=18 surviving daphnids in control groups 1-2, the mean and standard deviation are

$$\bar{Y}_0=162.5 \qquad \hat{\sigma}_0=57.835$$

Comparing the treatment group and control group averages by means of a (one tailed) two sample t-test we obtain

$$t = \frac{\bar{Y}_1 - \bar{Y}_0}{[(44\hat{\sigma}_1^2 + 17\hat{\sigma}_0^2)/61\ (1/45 + 1/18)]^{1/2}} = \frac{92.844 - 162.5}{13.32} = 5.23$$

with 61 degrees of freedom. This statistic is significant at $\alpha=0.0000$.

There is thus strong statistical evidence that beryllium diminishes reproduction. It appears that it is a substantial reduction. Since there is no response trend among the treatment groups, we conclude that the concentration at the lowest treatment group produces a (biologically and statistically) significant reduction in reproduction. This concentration is 2.60 mcg/l.

Goulden-Isophorone Test--Group 6 Deleted--Outlier in Group 1 Deleted

There are a control group and four treatment groups (2-5). As with the Chapman data, individual reproduction determinations are used for analysis. Responses only from daphnids that survived to the end of the test are included in the analysis. The outlying value (on the high side) in beaker 5 of group 1 is deleted. The previously discussed quadratic model was fitted to the data. The independent variable, v, was defined as v=(x-1.745)/0.443 (i.e., m=1.745, s=0.443). The results of the fit are shown in Figure XII.13 and the observed and predicted values are plotted vs $\log_{10}$ (1 + concentration) in Figure XII.14. The quadratic term is seen to be highly significant ($\hat{\beta}_2$=29.84 with estimated standard error 4.22) and strong quadratic trend is evident in Figure XII.14.

The output from CONFINT applied to this model shown in Figures XII.15, 16. Since the estimated average cumulative reproduction rate in the control group is 67.67 offspring per adult, the values of c corresponding to 10 percent and 20 percent reductions are -6.77 and -13.53 respectively. As in previous examples, the larger of the two roots in each output are the ones of biological importance. These correspond to roots no. 1 in the outputs. From the bottoms of Figures XII. 15, 16 we read these as

10 percent decrease: $\hat{conc}$=94.571 95 percent conf interval (77.730, 115.060)
20 percent decrease: $\hat{conc}$=103.136 95 percent conf interval (85.841, 123.915)

Due to the steepness of the response trend, both the concentrations associated with 10 percent and with 20 percent reductions in reproduction are reasonably precisely determined. However, because of this same steepness in trend, these two concentrations are close to one another and cannot be well separated by the information from this toxicity test.

C. LENGTH RESPONSES

The situation is essentially the same as for reproduction responses except for the need to adjust for the effects of beaker-to-beaker heterogeneity in those cases when the daphnids are multiply housed (e.g., LeBlanc's Tests A and B). These considerations have been discussed in detail in previous sections (see e.g., Subsection XI-C) and so need not be discussed again here. The models fitted and the notation used are the same as those for the reproduction response. We consider several examples.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted

There are four treatment groups and a control group. The concentrations corresponding to the two lowest treatment groups, Groups 3 and 4, are essentially the same. The quadratic model fitted is

$$Y=\mu+\beta_0 I+\beta_1 Iv+\beta_2 Iv^2+\varepsilon$$

where $v\equiv(x+2.257)/0.311$. The results of this fit are shown in Figure XII.17. The observed and predicted values are plotted in Figure XII.18 vs $\log_{10}$ (concentration). The quadratic term in Figure XII.17 has the wrong sign (positive), is small, and is nonsignificant. The display in Figure XII.18 shows no concentration related trend in average length within the range of the treatment groups. All the treatment groups appear to have substantially lower average length than the solvent control group.

The analysis shown in Figure XII.17 was carried out on a per beaker basis. That is, average lengths were calculated within each beaker and regression models were fitted to these averages. This of course resolves the issue of correlated responses within beakers. The residual mean square in Figure XII.17 corresponds conceptually to the mean square for beakers within groups that was discussed in subsection V-D. If the data had been completely balanced (i.e., equal numbers of daphnids per beaker) and if the same regression models had been fitted to the data (an analysis of variance model was used in Subsection V-D and a quadratic regression model was used in this analysis) then the two mean squares would be exactly the same. Since this data set is nearly balanced (except for two beakers in Group 6 with small numbers of daphnids), the two mean squares should be very close. This is in fact seen to be the case. The mean square for beakers within groups was calculated in Subsection VD to be MSI=0.2722. The average number of daphnids per beaker is N=16.7. When MSI is normalized to correspond to the variability of beaker averages, the estimated variance is approximately MSI/N=0.2722/16.7=0.0163. This is very similar to the error mean square value of 0.0176 calculated in Figure XII17. Thus the discussion in Subsection VD about augmenting the degrees of freedom of MSI based on pooling information from the variability among daphnids within beakers, also holds for this analysis. We might thus assign 98 degrees of freedom (i.e., the number calculated in Subsection V-D) to the error mean square rather than the 16 degrees of freedom obtained directly from the regression output. This would reduce the observed significance level for the F test for the quadratic coefficient from 0.265 to 0.251. If we used the mean square value 0.0163, from Subsection VD, this would further reduce the observed significance level to 0.233. Since such minor changes are of no practical importance, we do not pursue this possibility further.

The quadratic term was deleted from the model and the straight line model $Y=\mu+\beta_0 I+\beta_1 Ix+\varepsilon$ was fitted to the data. The results are shown in Figure XII.19. The linear term is essentially 0 and is nonsignificant ($\hat{\beta}_1=0.029$ with estimated standard error 0.111). We thus conclude that there is no significant trend in reproduction within the range of the treatment group concentrations. This is confirmed when we attempt to estimate concentrations resulting in specified reductions in average length

relative to the control group. As shown in Figure XII.20, the confidence interval on the concentration associated with a 10 percent reduction from the control group average ranges from 0 to infinity. It is thus meaningless.

Because of the lack of trend among the treatment group responses, we calculate an overall average length within the treatment groups for comparison with that in the control group. For the N=16 beakers in treatments groups 3-6, the mean and standard deviation of the beaker averages are

$$\bar{Y}_1=5.023 \qquad \hat{\sigma}_1=0.14032$$

For the N=4 beakers in the solvent control group, the mean and standard deviation of the beaker averages are

$$\bar{Y}_0=5.2775 \qquad \hat{\sigma}_0=0.5852$$

Comparing the treatment group and control group averages by means of a (one tailed) two sample t-test we obtain

$$t \equiv \frac{\bar{Y}_1 - \bar{Y}_0}{[(15\hat{\sigma}_1^2 + 3\hat{\sigma}_2^2)/18\ (1/16 + 1/4)]^{1/2}} = \frac{5.0231 - 5.2775}{0.0732} = -3.475$$

with 18 degrees of freedom. This statistic is significant at $\alpha=0.001$. If we had used the mean square calculated in Subsection VD, namely 0.2722/16.7 = 0.0163 with 98 degrees of freedom, then t would be equal to -3.565 with 98 degrees of freedom. This is significant at $\alpha=0.0003$. Thus the conclusions are unchanged for all practical purposes.

There is the strong statistical evidence that the toxicant in Test A diminishes average length in the treatment groups relative to that in the solvent control group. Whether the magnitude of decrease is of biological importance is a matter for biological judgement. Since there is no trend in response among the treatment groups, we conclude that the concentration corresponding to the lowest treatment group produces a statistically significant reduction in reproduction relative to the solvent control group. This concentration is 0.00290 mg/l.

Regression models fitted to these data with the water control group in place of the solvent control group would yield similar results except that the treatment group responses do not differ from the water control group response.

Chapman-Beryllium Test

The setup is the same as that for the reproduction responses. Since daphnids are individually housed in this test, there are no complications due to beaker-to-beaker heterogenity. The results of the analysis are shown in Figures XII.21 to XII.25.

The previously discussed quadratic model was fitted to the data. The independent variable, v, was defined as $v=(x-1.276)/0.551$. The results of the fit are shown in Figure XII.21 and the observed and predicted values are plotted in Figure XII.22 vs $\log_{10}$ (concentration). It is evident from both these displays that the trend within the treatment groups is linear, i.e., the quadratic component is essentially 0. The quadratic term was thus deleted from the model and the straight line model $Y=\mu+\beta_0 I+\beta_1 Ix+\varepsilon$ was fitted to the data. The results are shown in Figure XII.23. The linear term is very highly significant, as is evident from Figure XII.22.

The output from CONFINT applied to this model is shown in Figures XII.24,25. We calculate point and confidence interval estimates of concentrations corresponding to 10 percent and 20 reductions in average 21-day length, relative to the control group. Since the estimated average 21-day length in the control group is 4.278mm, the values of c specified are -0.428 and -0.856. Since we are dealing with a straight line model, there is just one root. From the bottoms of Figure II.24,25 we read these as

10 percent decrease: $\hat{conc}$=17.588 95 percent conf interval (5.10, 60.59)
20 percent decrease: $\hat{conc}$=359.205 95 percent conf interval (49.31, 2616.57)

The point estimate of the concentration corresponding to a 20 percent reduction in length exceeds the highest treatment group. It thus represents extrapolation beyond the range of the data, with the consequent danger of extrapolation bias. The confidence intervals are too wide to be useful. The ranges of concentrations in these intervals span factors of 12 and 53 respectively. We must therefore conclude that the results of this test do not provide enough information to precisely estimate the concentrations associated with 10 percent and with 20 percent reductions in average length relative to the control group.

```
BIONOMICS A&B, FERTILITY & LENGTH VS DOSE REGRESSIONS                    09/16/81     16.21.05.
TEST A, WATER CONTROL
FILE   NONAME   (CREATION DATE = 09/16/81 )
SUBFILE  TESTA      TEST3

* * * * * * * * * * * * * * * * * * * * MULTIPLE REGRESSION * * * * * * * * * * * * * * * * * * * *

DEPENDENT VARIABLE..    FERT21

MEAN RESPONSE     120.15000     STD. DEV.     27.05020

VARIABLE(S) ENTERED ON STEP NUMBER   1..   I
                                           IX
                                           IX2

MULTIPLE R            .58674      ANALYSIS OF VARIANCE    DF      SUM OF SQUARES      MEAN SQUARE          F      SIGNIFICANCE
R SQUARE              .34427      REGRESSION               3.        4786.17500       1595.39167       2.80005          .073
ADJUSTED R SQUARE     .22132      RESIDUAL                16.        9116.37500        569.77344
STD DEVIATION       23.86993      COEFF OF VARIABILITY    19.9 PCT

---------- VARIABLES IN THE EQUATION ----------                  ---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE          B        STD ERROR B        F                BETA          VARIABLE     PARTIAL   TOLERANCE        F
                                          SIGNIFICANCE       ELASTICITY                                          SIGNIFICANCE

I            41.293848      16.818355      6.0284223          .6264885
                                              .026              .27495
IX           -1.6552075      7.2957918     .51670695E-01      -.0546275
                                              .823             -.00001
IX2          -20.467907      10.899700      3.5262881         -.5367206
                                              .079              -.12804
(CONSTANT)   102.50000      11.934964      73.757387
                                              .000

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

I            282.85707
IX            40.04604   53.22858
IX2         -111.58378  -42.66969  118.80346

              I          IX         IX2
```

Figure XII.1. LeBlanc test A - water control group - group 7 deleted. Output from quadratic regression model of cumulative reproduction per surviving adult versus $\log_{10}$ (concentration).

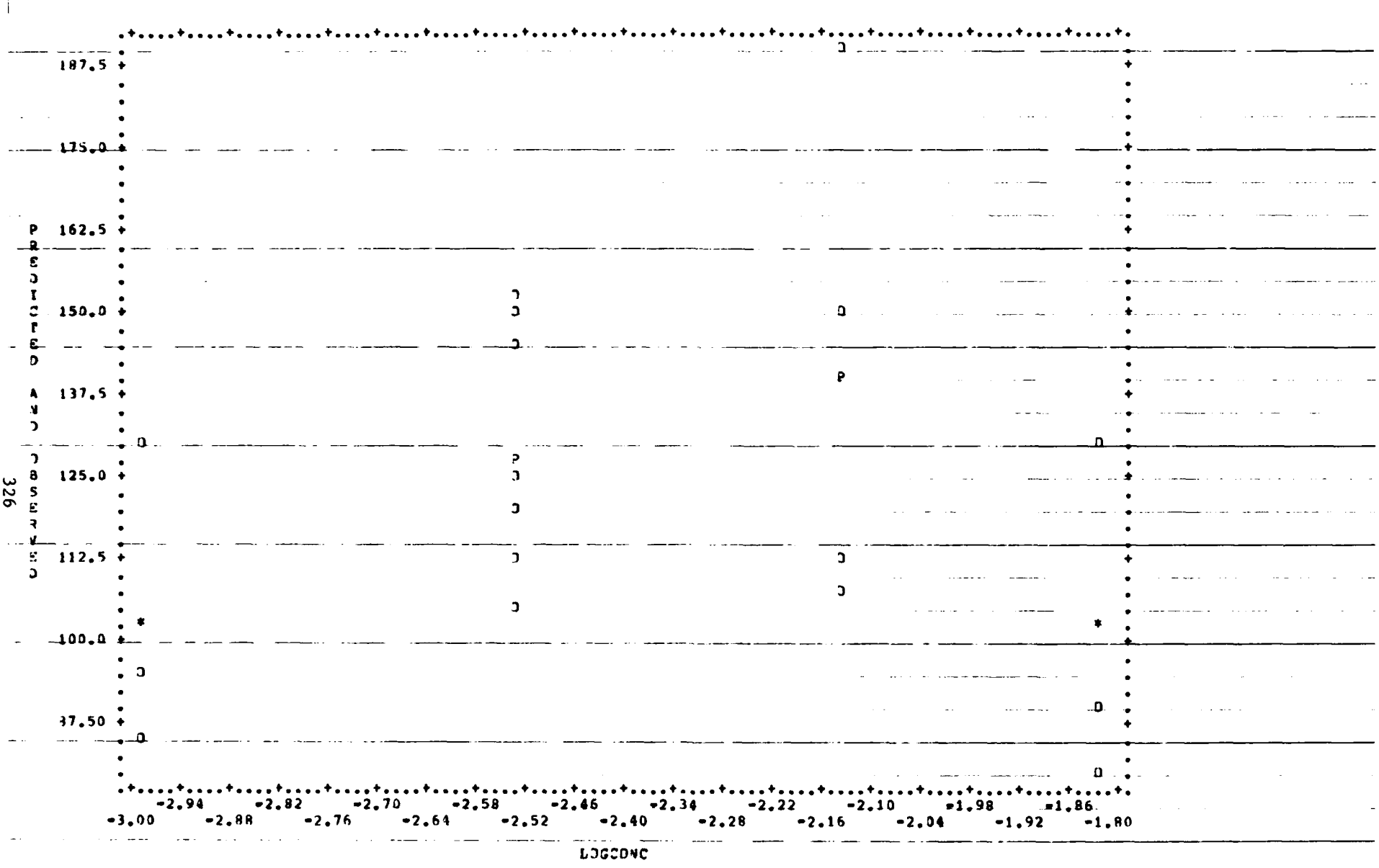


Figure XII.2. Observed and predicted values from quadratic regression model of cumulative reproduction per surviving adult versus $\log_{10}$ (concentration). LeBlanc test A water control group - group 7 deleted.

```
MINITAB TEST A, WATER CONTROL, REPRODUCTION VS. DOSE     (10% REDUCTION)

MODEL:  Y =   102.50000000 +    41.29385000I +    -1.65520000I V +   -20.46790000 I V**2   (QUADRATIC)

          WHERE I = 0 FOR CONTROL GROUP DAS.
                = 1 FOR TREATMENT GROUP DAS.

                V = ((LOGCONC -   -2.25700000)/     .31100000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION =   16

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

                    282.85737000   40.04604000 -111.58380000
                     40.04604000   53.22858000  -42.66969000
                   -111.58380000  -42.66969000  118.80346000

C =  -10.25000000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES   (ROOT 1)       .01538877     .02380626     .03682803
                      (ROOT 2)      -.01538877     .02505073    -.04077900

VARIANCE V =  (ROOT 1)=      .08632868
              (ROOT 2)=      .21420519

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.          POINT ESTIMATE          95% CONFIDENCE INTERVAL

RESCALED UNITS   (ROOT 1)    1.54698912    (     .92409618,    2.16988206)
                 (ROOT 2)   -1.62785721    (   -2.60904211,    -.64667231)

LOG-CONC. UNITS  (ROOT 1)   -1.77588638    (   -1.96960609,   -1.58216668)
                 (ROOT 2)   -2.76326359    (   -3.06841210,   -2.45811509)

RAW CONC. UNITS  (ROOT 1)     .01675381    (     .01072492,     .02617178)
                 (ROOT 2)     .00172479    (     .00085426,     .00348245)
```

Figure XII.3. Point and confidence interval estimates of concentration associated with 10 percent reduction in average reproduction relative to control group. LeBlanc test A - water control group - group 7 deleted.

BIOMETICS TEST A, WATER CONTROL, REPRODUCTION VS. DOSE (20% REDUCTION)

MODEL: Y = 132.50000000 + 41.29385000I + -1.65520000I V + -20.46790000 I V**2 (QUADRATIC)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = ((LOGCONC + 2.25700000)/ .31100000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 16

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

282.85707000	40.04604000	-111.58380000
40.04604000	53.22458000	-42.66969000
-111.58380000	-42.66969000	118.80346000

C = -20.50000000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES	(ROOT 1)	.01405540	.02386018	.04050459
	(ROOT 2)	-.01405540	.02499681	-.04445556

VARIANCE V (ROOT 1)= .09842753
(ROOT 2)= .25117928

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.		POINT ESTIMATE	95% CONFIDENCE INTERVAL	
RESCALED UNITS	(ROOT 1)	1.69758105	(1.03247000,	2.35269210)
	(ROOT 2)	-1.77844914	(-2.84094628,	-.71595200)
LOG-CONC. UNITS	(ROOT 1)	-1.72905229	(-1.93590183,	-1.52220276)
	(ROOT 2)	-2.8100976B	(-3.14053429,	-2.47966107)
RAW CONC. UNITS	(ROOT 1)	.01866155	(.01159039,	.03004673)
	(ROOT 2)	.00154847	(.00072355,	.00331390)

Figure XII.4. Point and confidence interval estimates of concentration associated with 20 percent reduction in average reproduction relative to control group. LeBlanc test A - water control group - group 7 deleted.

```
BIONOMICS A&B, FERTILITY & LENGTH VS DOSE REGRESSIONS                    09/16/81     16.21.05.
TEST A, SOLVENT CONTROL
FILE   NONAME   (CREATION DATE = 09/16/81 )
SUBFILE   TESTA      TESTB

* * * * * * * * * * * * * * * * * * * * MULTIPLE REGRESSION * * * * * * * * * * * * * * * * * * * *

DEPENDENT VARIABLE..    FERT21

MEAN RESPONSE     130.45000     STD. DEV.     27.67571

VARIABLE(S) ENTERED ON STEP NUMBER   1..   I
                                           IX
                                           IX2

MULTIPLE R              .64218     ANALYSIS OF VARIANCE    DF      SUM OF SQUARES      MEAN SQUARE      F      SIGNIFICANCE
R SQUARE                .41240     REGRESSION               3.        6001.57500       2000.52500     3.74307      .033
ADJUSTED R SQUARE       .30222     RESIDUAL                16.        8551.37500        534.46094
STD DEVIATION         23.11841     COEFF OF VARIABILITY    17.7 PCT

---------- VARIABLES IN THE EQUATION ----------                    ---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE          B          STD ERROR B          F                 BETA          VARIABLE     PARTIAL   TOLERANCE      F
                                                SIGNIFICANCE      ELASTICITY                                          SIGNIFICANCE

I           =10.206152       16.288849          .39259363        =.1513427
                                                    .540             =.06259
IX           =1.6552075       7.0660925         .54871428E-01     =.0531973
                                                    .818             =.00001
IX2         =20.467907       10.556536          3.7592743        =.5245899
                                                    .070             =.11793
(CONSTANT)   154.00000       11.559206         177.49473
                                                    .000

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

I          265.32661
IX          37.56414   49.92966
IX2       =104.66823  =40.02518  111.44045

              I          IX         IX2
```

Figure XII.5. LeBlanc test A - solvent control group - group 7 deleted. Output from quadratic regression of cumulative reproduction per surviving adult versus $\log_{10}$ (concentration).

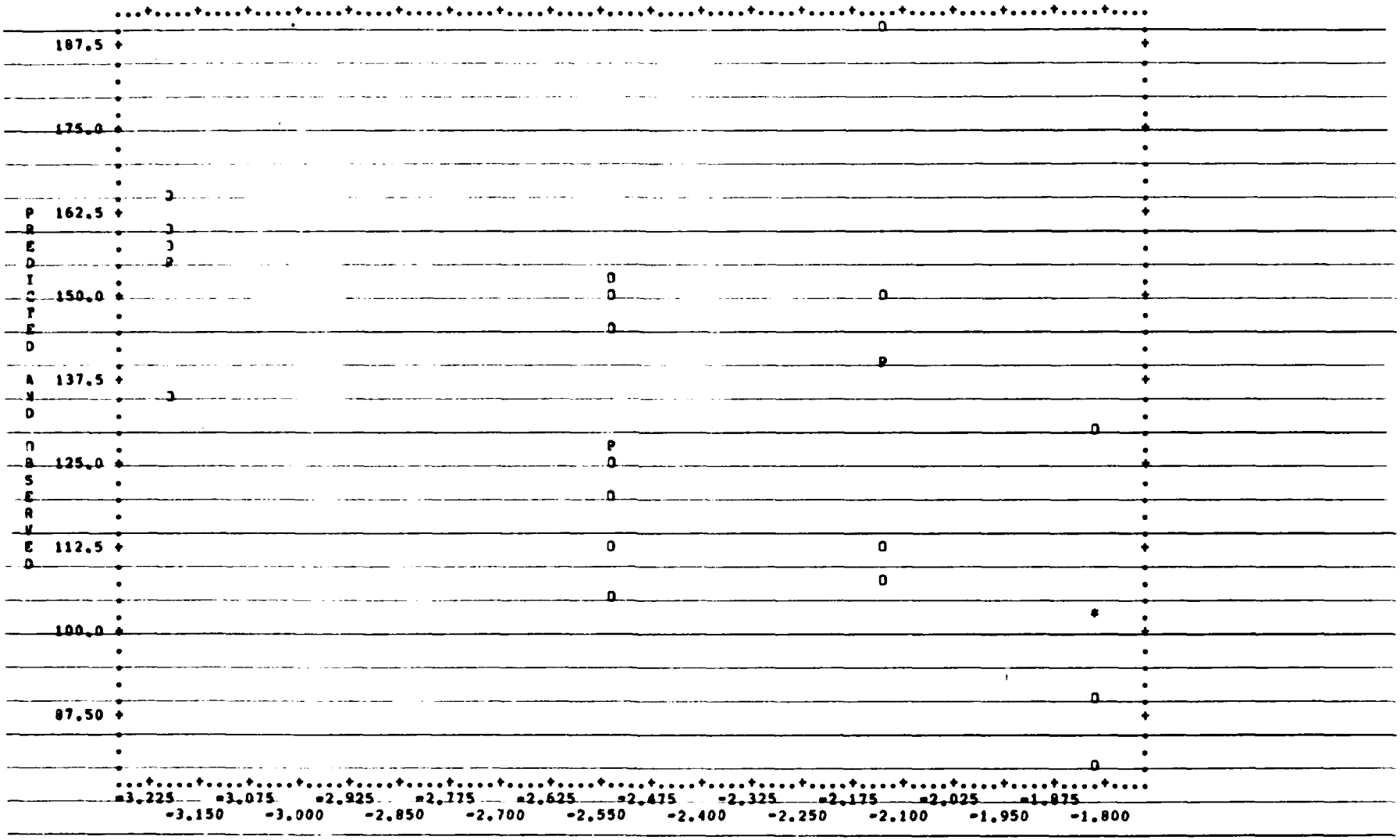


Figure XII.6. Observed and predicted values from quadratic regression model of cumulative reproduction per surviving adult versus $\log_{10}$ (concentration). LeBlanc test A - solvent control group - group 7 deleted.

1 BIONOMICS TEST A, SOLVENT CONTROL, REPRODUCTION VS DOSE (10% REDUCTION)

MODEL: Y = 154.00000000 + -10.20615000I + -1.65521000I V + -20.46790000 I V**2 (QUADRATIC)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = ((LOGCONC - -2.25200000)/ .31100000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 16

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

265.32660000	37.56414000	-104.66820000
37.56414000	49.92956000	-40.02520000
-104.66820000	-40.02520000	111.44065000

C = -15.40000000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES	(ROOT 1)	.04833860	.02247396	.01044877
	(ROOT 2)	-.04833860	.02638303	-.01439976

VARIANCE V (ROOT 1)= .61443978
(ROOT 2)= .46671711

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.		POINT ESTIMATE	95% CONFIDENCE INTERVAL
RESCALED UNITS	(ROOT 1)	.46492777	(-1.19685986, 2.12671560)
	(ROOT 2)	-.54579635	(-1.99411032, .90251762)
LOG-CONC. UNITS	(ROOT 1)	-2.11240746	(-2.62922342, -1.59559151)
	(ROOT 2)	-2.42674267	(-2.87716831, -1.97631702)
RAW CONC. UNITS	(ROOT 1)	.00771956	(.00234842, .02537514)
	(ROOT 2)	.00374332	(.00132688, .01056046)

Figure XII.7. Point and confidence interval estimated of concentration associated with 10 percent reduction in average reproduction relative to control group. LeBlanc test A - solvent control group - group 7 deleted.

BIOMETICS TEST A, SOLVENT CONTROL, REPRODUCTION VS DOSE (20% REDUCTION)

MODEL: Y = 154.00000000 + -10.20615000I + -1.65521000I V + -20.46790000 I V**2 (QUADRATIC)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = ((LOGCONC - -2.25700000)/ .31100000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 16

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

265.32660000	37.56414000	-104.66820000
37.56414000	49.92966000	-40.02520000
-104.66820000	-40.02520000	111.44045000

C = -30.80000000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES	(ROOT 1)	.02433392	.02344457	.02258773
	(ROOT 2)	-.02433392	.02541242	-.02653872

VARIANCE V (ROOT 1)= .12681941
(ROOT 2)= .14018379

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.		POINT ESTIMATE	95% CONFIDENCE INTERVAL	
RESCALED UNITS	(ROOT 1)	.96345239	(.20848409,	1.71842070)
	(ROOT 2)	-1.04432097	(-1.83807283,	-.25056911)
LOG-CONC. UNITS	(ROOT 1)	-1.95736631	(-2.19216145,	-1.72257116)
	(ROOT 2)	-2.58178382	(-2.82864065,	-2.33492699)
RAW CONC. UNITS	(ROOT 1)	.01103148	(.00642449,	.01894213)
	(ROOT 2)	.00261949	(.00148375,	.00462459)

Figure XII.8. Point and confidence interval estimates of concentration associated with 20 percent reduction in average reproduction relative to control group. LeBlanc test A - solvent control group - group 7 deleted.

CHAPMAN TEST- FERTILITY & LENGTH VS DOSE REGRESSIONS 09/16/81 16.21.15.

FILE NONAME (CREATION DATE = 09/16/81)

********** MULTIPLE REGRESSION **********

DEPENDENT VARIABLE.. TOTALYNG

MEAN RESPONSE 112.74603 STD. DEV. 57.01159

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX
IX2

MULTIPLE R	.56368	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.31773	REGRESSION	3.	64028.95332	21342.98444	9.15868 .000
ADJUSTED R SQUARE	.28304	RESIDUAL	59.	137490.98319	2330.35565	
STD DEVIATION	48.27376	COEFF OF VARIABILITY	42.8 PCT			

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION			
VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY	VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE
I	-63.757279	15.779618	16.325504 .000	-.5092640 -.40392				
IX	1.5199769	7.6538962	.39437450E-01 .843	.0224405 .00000				
IX2	-6.0421796	8.4330062	.51336047 .477	-.0936717 -.03736				
(CONSTANT)	162.50000	11.378234	203.96565 0					

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

I	248.99634		
IX	-19.33777	58.58213	
IX2	-69.41058	19.82906	71.11559
	I	IX	IX2

Figure XII.9. Chapman-Beryllium test--combined control groups. Output from quadratic regression model of cumulative reproduction versus $\log_{10}$ (concentration). Only surviving daphnids are included.

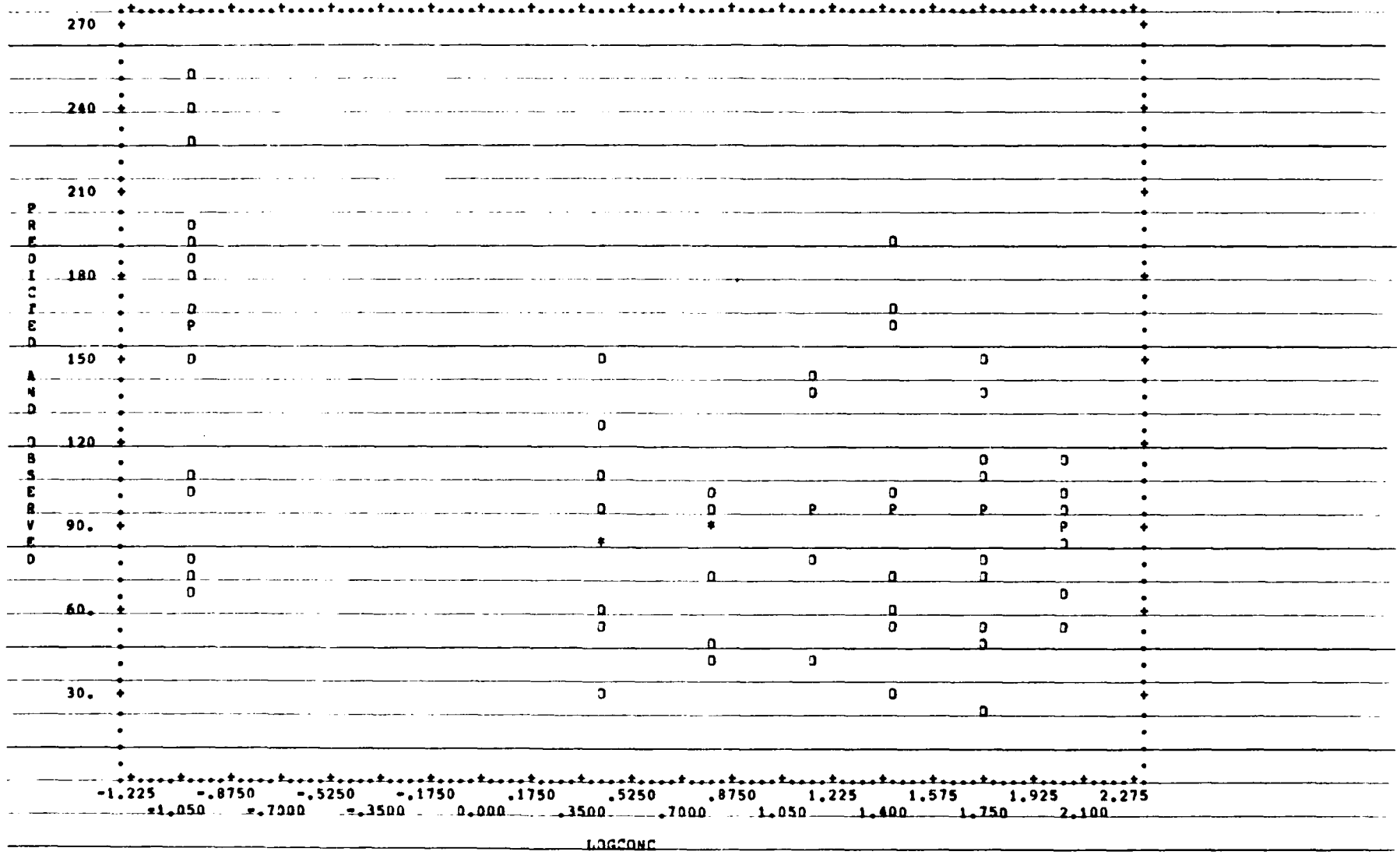


Figure XII.10. Observed and predicted values from quadratic regression model of cumulative reproduction versus $\log_{10}$ (concentration). Surviving adults only. Chapman-Beryllium-combined control groups.

CHAPMAN TEST- FERTILITY & LENGTH VS DOSE REGRESSIONS 09/25/81 11.39.51.

FILE NONAME (CREATION DATE = 09/25/81)

* MULTIPLE REGRESSION *

DEPENDENT VARIABLE.. TOTALYNG

MEAN RESPONSE 112.74603 STD. DEV. 57.01159

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX

MULTIPLE R	.55838	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.31179	REGRESSION	2.	62832.64085	31416.32042	13.59158 .000
ADJUSTED R SQUARE	.28885	RESIDUAL	60.	138687.29566	2311.45493	
STD DEVIATION	48.07759	COEFF OF VARIABILITY	42.6 PCT			

VARIABLES IN THE EQUATION

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
I	-77.076028	21.492262	12.860990 .001	-.6156481 -.48830
IX	5.8161690	13.165455	.19516515 .660	.0758397 .04701
(CONSTANT)	162.50000	11.331997	205.63347 0	

VARIABLES NOT IN THE EQUATION

VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

	I	IX
I	461.91734	
IX	-221.13949	173.32921

Figure XII.11. Chapman-beryllium test-combined control groups. Output from linear regression model of cumulative reproduction versus $\log_{10}$ (concentration). Only surviving daphnids are included.

FARMAN TEST, REPRODUCTION VS DOSE (10% REDUCTION)

MODEL: Y = 162.50000000 + -77.07603000I + 5.81616900I V (LINEAR)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = LOG(CONC)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 60

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

461.91734000 -221.13949000
-221.13949000 173.32921000

C = -16.25000000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES -.17193448 -1.79810661

VARIANCE V = 437.32733390

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.	POINT ESTIMATE	95% CONFIDENCE INTERVAL

GOLDEN TEST- FERTILITY VS DOSE REGRESSIONS 09/16/81 16.21.24.

FILE NONAME (CREATION DATE = 09/16/81)

********** MULTIPLE REGRESSION **********

DEPENDENT VARIABLE.. TOTALYNG

MEAN RESPONSE 64.81818 STD. DEV. 31.44882

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX
IX2

MULTIPLE R	.88380	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.78110	REGRESSION	3.	24721.05329	8240.35110	34.49410 0
ADJUSTED R SQUARE	.75846	RESIDUAL	29.	6927.85580	238.89158	
STD DEVIATION	15.45612	COEFF OF VARIABILITY	23.8 PCT			

VARIABLES IN THE EQUATION

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
I	25.285484	8.0753165	9.8044544 .004	.3149143 .31917
IX	-45.989199	4.5374133	102.72946 0	-1.3179584 -.00051
IX2	-29.840168	4.2224695	49.942407 0	-.9861340 -.36261
(CONSTANT)	67.666667	6.3099337	115.00057 0	

VARIABLES NOT IN THE EQUATION

VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

	I	IX	IX2
I	65.21074		
IX	-13.74117	20.58812	
IX2	-17.17649	14.25506	17.82925

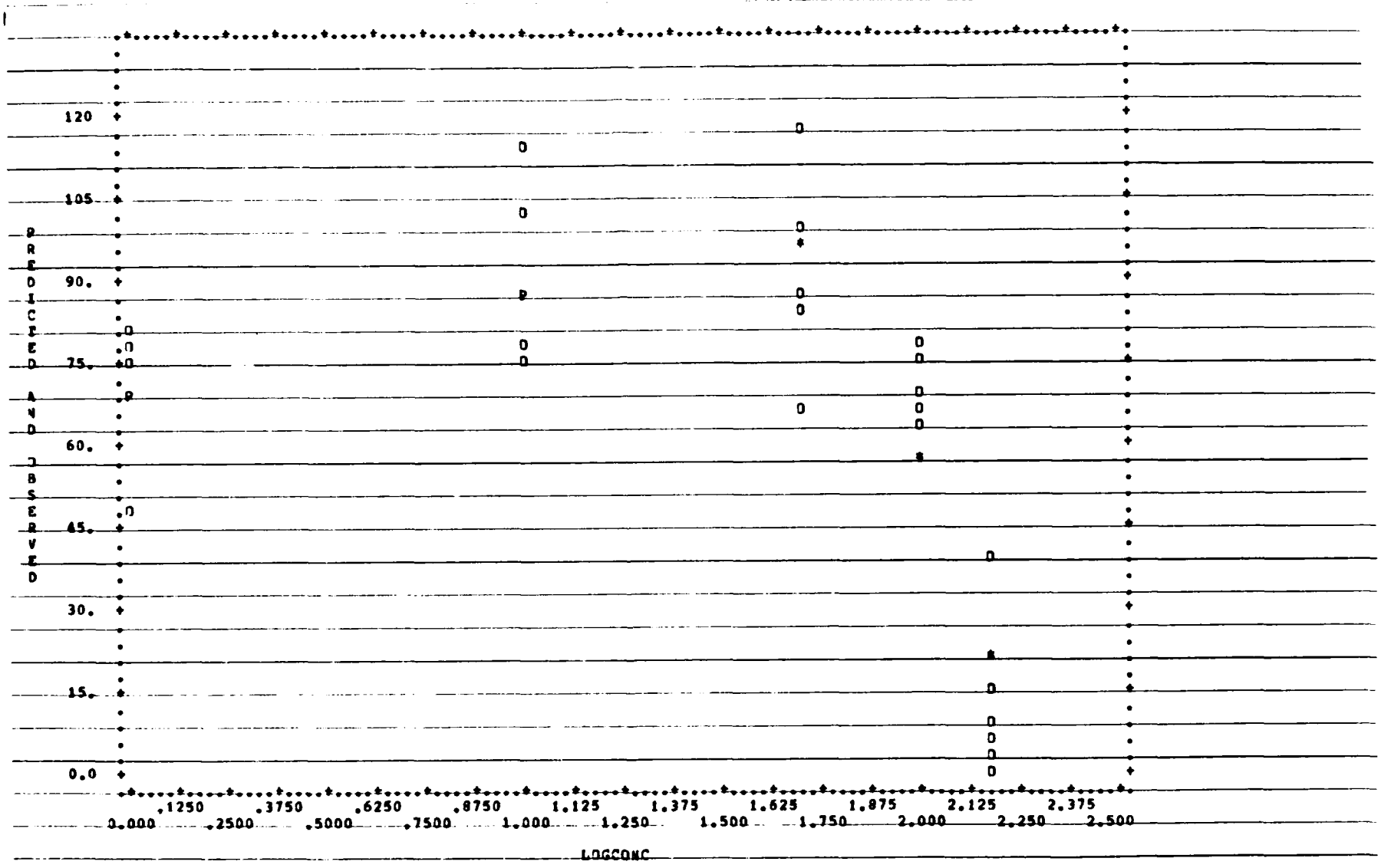


Figure XII.14. Observed and predicted values from quadratic regression model of cumulative reproduction versus $\log_{10}$ (concentration). Surviving adults only. Goulden isophorone - group 6 deleted - outlier in group 1 deleted.

GOULDEN TEST, REPRODUCTION VS DOSE (OMITTING OUTLIER) (10% REDUCTION)

MODEL: Y = 67.66667000 + 25.28548400I + -45.98920000I V + -29.84016800 I V**2 (QUADRATIC)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = ((LOGCONC = 1.74500000)/ .44300000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 29

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

65.21074000	-13.74117000	-17.17649000
-13.74117000	20.58812000	14.25506000
-17.17649000	14.25506000	17.82925000

C = -6.76666700 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES	(ROOT 1)	.01297414	.00675817	.00352030
	(ROOT 2)	-.01297414	.02675371	-.05516828

VARIANCE V (ROOT 1)= .00883767
(ROOT 2)= .02284809

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.		POINT ESTIMATE	95% CONFIDENCE INTERVAL
RESCALED UNITS	(ROOT 1)	.52089534	(.32864721, .71314347)
	(ROOT 2)	-2.06207969	(-2.37119339, -1.75296599)
LOG-CONC. UNITS	(ROOT 1)	1.97575664	(1.89059071, 2.06092256)
	(ROOT 2)	.83149870	(.69456133, .96843607)
RAW CONC. UNITS	(ROOT 1)	94.57070684	(77.73036593, 115.05951999)
	(ROOT 2)	6.78420085	(4.94949999, 9.29899610)

Figure XII.15. Point and confidence interval estimates of concentrations associated with 10 percent reduction in average reproduction relative to control group - Goulden isophorone - group 6 deleted - outlier in group 1 deleted.

```
GOULDEN TEST, REPRODUCTION VS DOSE (OMITTING OUTLIER)      (20% REDUCTION)

MODEL:  Y =    67.66667000 +    25.28548400I +   -45.98920000I V +   -29.84016800 I V**2   (QUADRATIC)

          WHERE I = 0 FOR CONTROL GROUP OBS.
                = 1 FOR TREATMENT GROUP OBS.

                V = ((LOGCONC =    1.74500000)/     .44300000)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION =   29

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

                     65.21074000  -13.74117000  -17.17649000
                    -13.74117000   20.58812000   14.25506000
                    -17.17649000   14.25506000   17.82925000

C =  -13.53333400 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES   (ROOT 1)        .01217301     .00737551     .00446875
                      (ROOT 2)       -.01217301     .02613637    -.05611673

VARIANCE V    (ROOT 1)=      .00774259
              (ROOT 2)=      .02333433

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.          POINT ESTIMATE        95% CONFIDENCE INTERVAL

RESCALED UNITS   (ROOT 1)     .60589019     (     .42594662,     .78583376)
                 (ROOT 2)   -2.14707454     (   -2.45946012,   -1.83468897)

LOG-CONC. UNITS  (ROOT 1)    2.01340935     (    1.93369435,    2.09312435)
                 (ROOT 2)     .79384598     (     .65545917,     .93223279)

RAW CONC. UNITS  (ROOT 1)  103.13577922     (   85.84091820,  123.91513487)
                 (ROOT 2)    6.22079626     (    4.52333933,    8.55525162)
```

Figure XII.16. Point and confidence interval estimates of concentrations associated with 20 percent reduction in average reproduction relative to the control group - Goulden isophorone - group 6 deleted - outlier in group 1 deleted.

```
BIOMETICS A&B, FERTILITY & LENGTH VS DOSE REGRESSIONS                    09/16/81      16.21.05.
TEST A, SOLVENT CONTROL
FILE    NONAME    (CREATION DATE =  09/16/81 )
SUBFILE   TESTA       TESTB

* * * * * * * * * * * * * * * * * * * * * * * M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * * * * * * * * * * *

DEPENDENT VARIABLE..    LENGTH21

MEAN RESPONSE        5.07400      STD. DEV.          .16427

VARIABLE(S) ENTERED ON STEP NUMBER   1..   I
                                          IX
                                          IX2

MULTIPLE R            .67236      ANALYSIS OF VARIANCE    DF      SUM OF SQUARES      MEAN SQUARE          F      SIGNIFICANCE
R SQUARE              .45207      REGRESSION               3.          .23177            .07726          4.40028        .019
ADJUSTED R SQUARE     .34933      RESIDUAL                16.          .28091            .01756
STD DEVIATION         .13250      COEFF OF VARIABILITY    2.6 PCT

---------- VARIABLES IN THE EQUATION ----------          ---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE          B          STD ERROR B        F            BETA          VARIABLE    PARTIAL   TOLERANCE        F
                                              SIGNIFICANCE   ELASTICITY                                       SIGNIFICANCE

I            -.32004776     .93359381E-01    11.752051     -.7995885
                                                .003         -.05046
IX           -.15941109E-01  .40499240E-01     .15493278    -.0863194
                                                .699         -.00000
IX2           .69914022E-01  .60504683E-01    1.3352131      .3019000
                                                .265          .01036
(CONSTANT)   5.2775000       .66251474E-01  6345.4933
                                                .000

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

I             .00872
IX            .00123     .00164
IX2          -.00344    -.00131     .00366

              I          IX         IX2
```

Figure XII.17. LeBlanc test A - solvent control group - group 7 deleted. Output from quadratic regression model of 21 day lengths of surviving adults versus $\log_{10}$ (concentration).

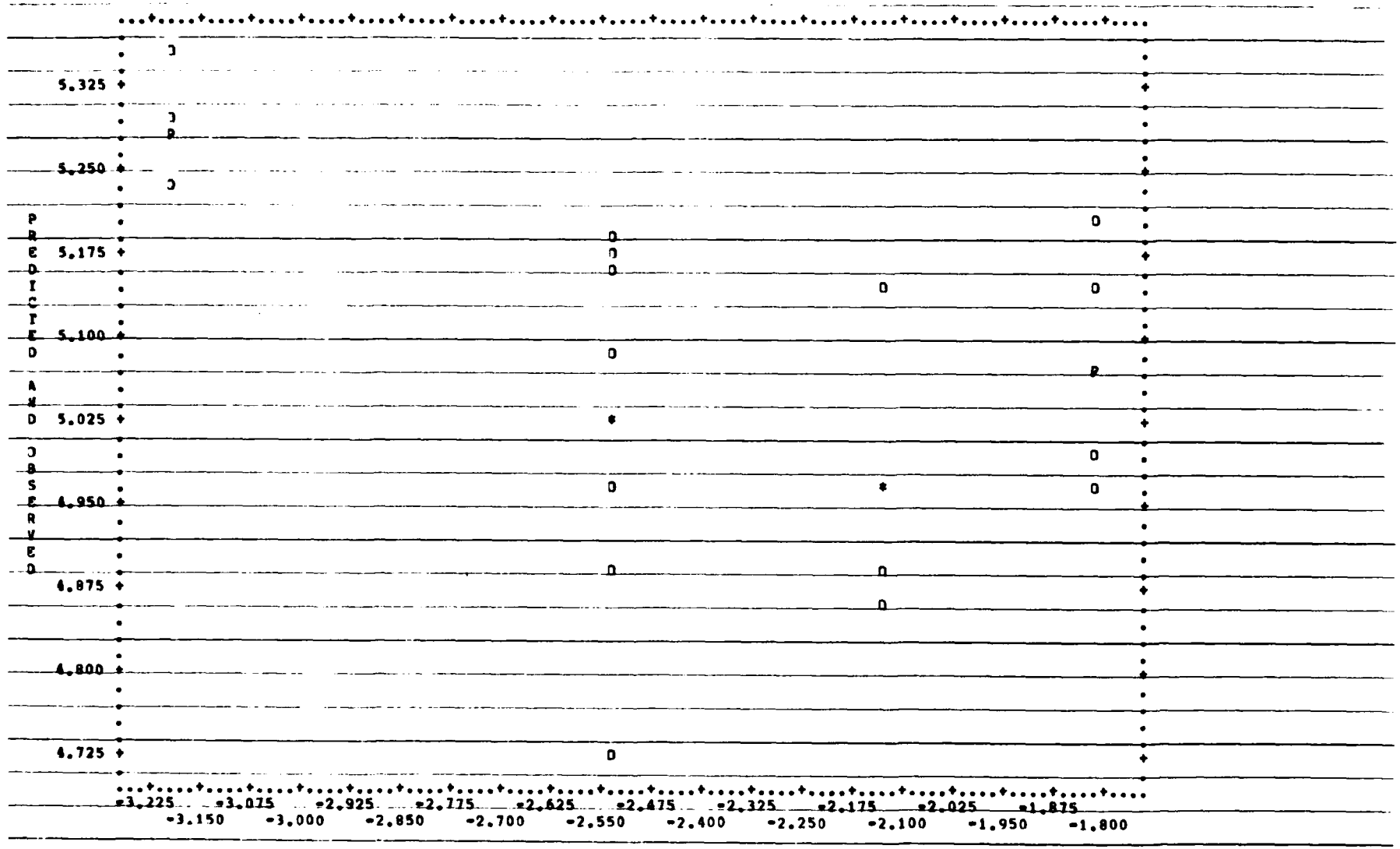


Figure XII.18. Observed and predicted values from quadratic regression model of 21 day lengths of surviving adults versus $\log_{10}$ (concentration). LeBlanc test A - solvent control group - group 7 deleted.

BIOMETICS ASA, FERTILITY & LENGTH VS DOSE REGRESSIONS 09/25/81 11.40.01.
SOLVENT CONTROL
FILE NONAME (CREATION DATE = 09/25/81)
SUBFILE TESTA TESTA

* M U L T I P L E R E G R E S S I O N *

DEPENDENT VARIABLE.. LENGTH21

MEAN RESPONSE 5.07400 STD. DEV. .16427

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX

MULTIPLE R	.63745	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.40635	REGRESSION	2.	.20833	.10416	5.81809 .012
ADJUSTED R SQUARE	.33650	RESIDUAL	17.	.30435	.01790	
STD DEVIATION	.13380	COEFF OF VARIABILITY	2.6 PCT			

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION			
VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY	VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE
I	-.18783827	.26135578	.51653999 .482	-.4692842 -.02962				
IX	.29483443E-01	.11096663	.70594623E-01 .794	.1734881 -.01049				
(CONSTANT)	5.2775000	.66901473E-01	6222.7898 0					

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

	I	IX
I	.06831	
IX	-.02779	.01231

Figure XII.19. LeBlanc test A - solvent control group - group 7 deleted. Output of linear regression model of 21 day lengths of surviving adults versus $\log_{10}$ (concentration).

BIONOMICS TEST A, SOLVENT CONTROL, LENGTH VS DOSE (10% REDUCTION)

MODEL: Y = 5.27750000 + -.18783830I + .02948344I V (LINEAR)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = LOG(CONC)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 17

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

.06831000 .02779000
.02779000 .01231000

C = -.52775000 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES =33.91734479 391.02975522

VARIANCE V = 1223.69407630

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.	POINT ESTIMATE	95% CONFIDENCE INTERVAL
LOG-CONC. UNITS	-11.52890233	(-85.33952756, 62.28172291)
RAW CONC. UNITS	.00000000	(.00000000,**********)

Figure XII.20. Point and confidence interval estimates of concentration associated with 10 percent reduction in average 21 day length relative to control group. LeBlanc test A - solvent control group - group 7 deleted.

CHAPM N TEST- FERTILITY & LENGTH VS DOSE REGRESSIONS 09/16/81 16.21.15.

FILE NONAME (CREATION DATE = 09/16/81)

* MULTIPLE REGRESSION *

DEPENDENT VARIABLE.. LENGTH

MEAN RESPONSE 3.96508 STD. DEV. .40288

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX
IX2

MULTIPLE R	.62167	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.38622	REGRESSION	3.	3.88659	1.29553	12.37517 .000
ADJUSTED R SQUARE	.35501	RESIDUAL	59.	6.17658	.10469	
STD DEVIATION	.32356	COEFF OF VARIABILITY	8.2 PCT			

---------- VARIABLES IN THE EQUATION ----------

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
I	-.44261830	.10576295	17.514255 .000	-.5003041 -.07974
IX	-.17854233	.51300267E-01	12.112761 .001	-.3730173 .00001
IX2	.49042474E-02	.56522254E-01	.75284537E-02 .931	.0107592 .00086
(CONSTANT)	4.2777778	.76262652E-01	3146.3906 0	

---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

	I	IX	IX2
I	.01119		
IX	-.00087	.00263	
IX2	-.00312	.00089	.00319

Figure XII.21. Chapman beryllium test - combined control groups. Output from quadratic regression model of 21 day lengths of surviving adults versus $\log_{10}$ (concentration).

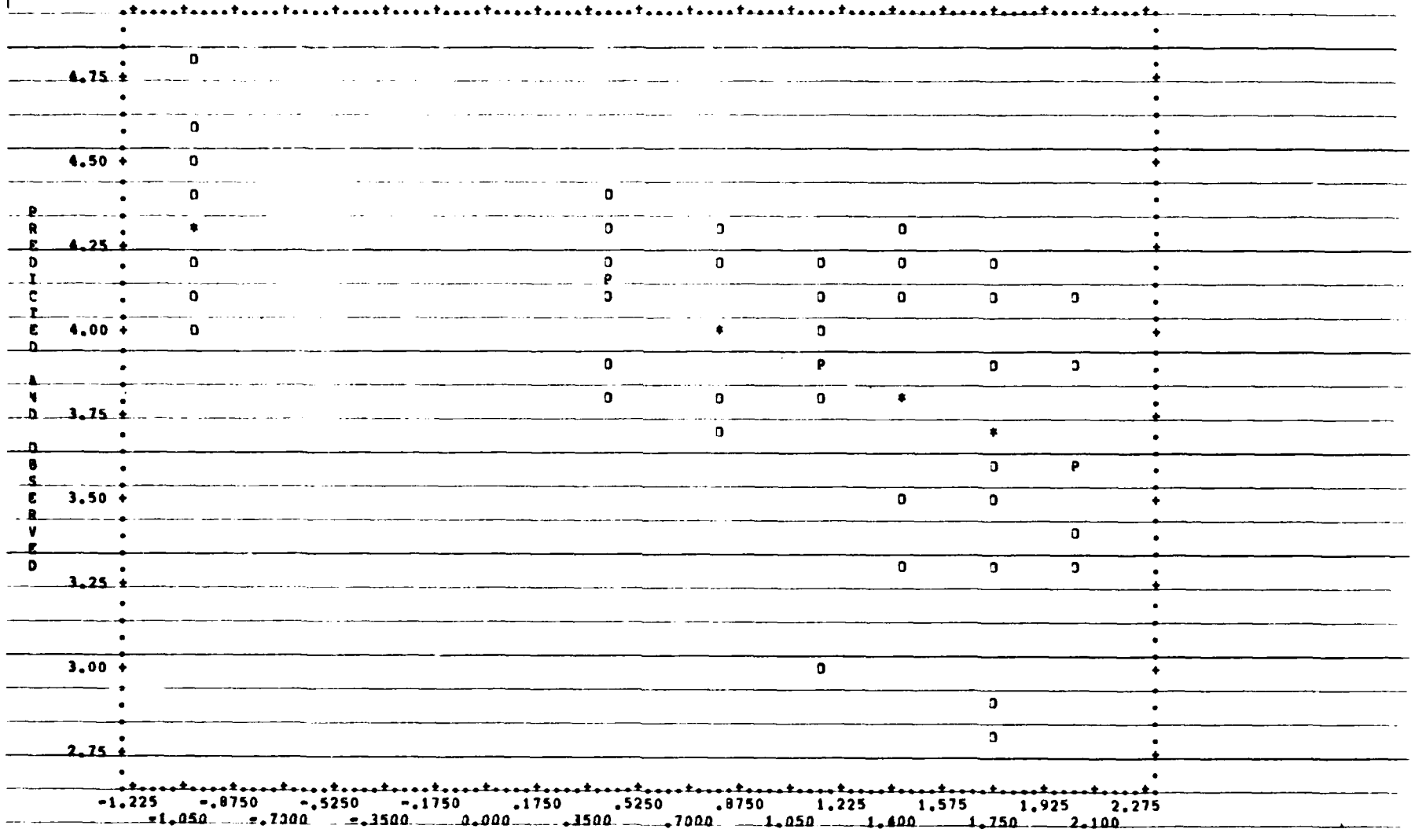


Figure XII.22. Observed and predicted values from quadratic regression model of 21 day lengths versus $\log_{10}$ (concentration). Surviving adults only. Chapman beryllium - combined control groups.

CHAPMAN TEST- FERTILITY & LENGTH VS DOSE REGRESSIONS 09/25/81 11.39.51.

FILE NONAME (CREATION DATE = 09/25/81)

* M U L T I P L E R E G R E S S I O N *

DEPENDENT VARIABLE.. LENGTH

MEAN RESPONSE 3.96508 STD. DEV. .40288

VARIABLE(S) ENTERED ON STEP NUMBER 1.. I
IX

MULTIPLE R	.62140	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F SIGNIFICANCE
R SQUARE	.38614	REGRESSION	2.	3.88580	1.94290	18.87115 .000
ADJUSTED R SQUARE	.36558	RESIDUAL	60.	6.17737	.10296	
STD DEVIATION	.32087	COEFF OF VARIABILITY	8.1 PCT			

VARIABLES IN THE EQUATION

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
I	.21198475E-01	.14343848	.21841272E-01	.0239612
			.883	-.00382
IX	-.32651501	.87865708E-01	13.809167	-.6024963
			.000	-.07504
(CONSTANT)	4.2777778	.75629285E-01	3199.3110	
			0	

VARIABLES NOT IN THE EQUATION

VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE

ALL VARIABLES ARE IN THE EQUATION.

VARIANCE/COVARIANCE MATRIX OF THE UNNORMALIZED REGRESSION COEFFICIENTS.

	I	IX
I	.02057	
IX	-.00985	.00772

Figure XII.23. Chapman beryllium test - combined control groups. Output from linear regression model of 21 day lengths versus $\log_{10}$ (concentration). Only surviving daphnids are included.

CHAPMAN TEST, LENGTH VS DOSE (10% REDUCTION)

MODEL: Y = 4.27778000 + -.02119848I + -.32651501I V (LINEAR)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = LOG(CONC)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 60

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

.02057000	-.00985000
-.00985000	.00772000

C = -.42777800 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES 3.06264634 3.81363564

VARIANCE V = .07512841

SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

INDEPENDENT VAR.	POINT ESTIMATE	95% CONFIDENCE INTERVAL	
LOG-CONC. UNITS	1.24520928	(.70798187,	1.78243668)
RAW CONC. UNITS	17.58770929	(5.10483695,	60.59498495)

Figure XII.24. Point and confidence interval estimates of concentration associated with 10 percent reduction in average 21 day length relative to control group. Chapman beryllium test - combined control groups.

CHAPMAN TEST, LENGTH VS DOSE (20% REDUCTION)

MODEL: Y = 4.27778000 + -.02119848I + -.32651501I V (LINEAR)

WHERE I = 0 FOR CONTROL GROUP OBS.
= 1 FOR TREATMENT GROUP OBS.

V = LOG(CONC)

DEGREES OF FREEDOM FOR ERROR IN REGRESSION = 60

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS (EXCLUDING U)

.02057000	-.00985000
-.00985000	.00772000

C = -.45555600 = PROPOSED CHANGE W.R.T. CONTROL MEAN U

PARTIAL DERIVATIVES 3.06264634 7.82610883

VARIANCE V = .19359545

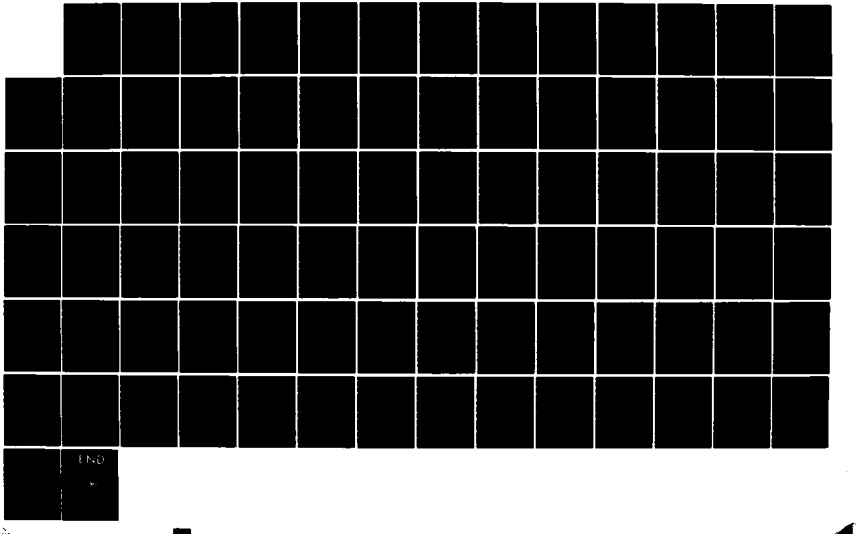
SOLUTION FOR VALUE(S) OF INDEPENDENT VAR. PRODUCING C UNITS CHANGE IN Y

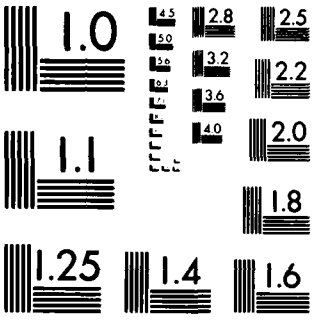
INDEPENDENT VAR.	POINT ESTIMATE	95% CONFIDENCE INTERVAL
LOG-CONC. UNITS	2.55534200	(1.69295214, 3.41773187)
RAW CONC. UNITS	359.20469462	(49.31194545, 2616.56707010)

Figure XII.25. Point and confidence interval estimates of concentration associated with 20 percent reduction in average 21 day length relative to control group. Chapman beryllium test - combined control groups.

AD-A152 070 DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY WITH DAPHNIA MAGNA(U) BATTELLE COLUMBUS LABS OH P I FEDER DEC 81 BATT-G-7673 DAMD17-80-C-0165 5/5

UNCLASSIFIED F/G 6/20 NL





XIII. CONFIDENCE INTERVAL PROCEDURES FOR COMPARISON OF EFFECT LEVELS BETWEEN TREATMENT AND CONTROL GROUPS

A. INTRODUCTION

In Sections IX-XII we considered the comparisons of effect levels in the treatment groups with those in the control groups for survival, reproduction, and length responses. We utilized hypothesis testing, multiple comparisons, and dose response estimation procedures. In this section we construct confidence intervals to quantify the extent of the differences between the treatment group and control group effect levels. The values contained within the confidence intervals indicate the extent of biological significance of these effect differences. The widths of the confidence intervals indicate the degree of precision in the data for estimating these differences. Narrow confidence intervals signify precise estimates while wide confidence intervals signify imprecise estimates. By contrast, hypothesis testing procedures merely state whether the null hypothesis was accepted or rejected; they give no indication of the extent of the effect.

Confidence interval comparisons of treatment group and control group average responses are also a very useful adjunct to hypothesis testing procedures, particularly at the MATC. After an MATC has been determined, it is worthwhile to calculate a confidence interval there. The upper confidence bound (for mortality) or the lower confidence bound (for length or reproduction) indicate how much worse than the control group, the MATC could conceivably be. If this confidence bound is biologically very undesirable, then the MATC might be too high a concentration even though it is not statistically significantly different than the control. Thus it might make sense to be more conservative and report a concentration lower than the MATC. Such a phenomenon will occur most frequently if the hypothesis testing procedure has very poor power. Thus the calculation and utilization of confidence intervals at the MATC can help to alleviate one of the principal weakness of the hypothesis testing approach to determining safe concentrations.

Confidence interval procedures are discussed in a number of places. Feder and Collins [1], Section XIII, discusses a number of approaches for placing confidence intervals on the ratios of treatment group mortality rates to control group rates in fathead minnow early life stage tests. In this section we use some of these methods, as well as others, to place confidence intervals on pairwise treatment-control ratios and differences in survival, length, and reproduction effects observed in 21-day chronic Daphnia tests. We account for possible beaker-to-beaker heterogeneity of effects within groups by the adjustment techniques discussed in Section V.

We illustrate confidence interval calculations using unsmoothed effect levels (i.e., using average observed effect levels within each treatment group, unadjusted for those in the other groups) and using smoothed effect levels (i.e., using effect levels based on the predictions from regression

models fitted to all the treatment groups). We illustrate these approaches for each of survival, length, and reproduction.

B. MORTALITY--UNSMOOTHED RESPONSES

Feder and Collins [1] discusses three approaches to the construction of confidence intervals for mortality data. These are based on large sample normal theory, exact small sample theory based on the noncentral conditional distribution of the 2 x 2 contingency table, and on Poisson theory, (most appropriate when the response probabilities are small). We illustrate here the asymptotic approach, the Poisson approach, and a variant on the exact approach due to Thomas and Gart [30]. We illustrate these procedures with 21-day cumulative mortality data from several of the examples considered previously.

Asymptotic Approach

Let p_c, p_t denote the (population) mortality probabilities in the control group and in a treatment group respectively and let $q=1-p$. Let N_c, N_t denote the associated (effective) sample sizes in these groups, let $\phi \equiv \ell n(p_t/p_c)$, and let p_t, p_c, ϕ denote estimates of these quantities. Feder and Collins [1], Subsection XIIIB, state that an asymptotic 95% confidence interval on ϕ is

$$\hat{\phi} - 1.96\left[\frac{q_c}{N_c p_c} + \frac{q_t}{N_t p_t}\right]^{1/2} \leq \phi \leq \hat{\phi} + 1.96\left[\frac{q_c}{N_c p_c} + \frac{q_t}{N_t p_t}\right]^{1/2}$$

This interval is valid as $N_c, N_t \to \infty$ with p_c, p_t fixed.

In the case of LeBlanc's Test A the effective sample sizes and observed mortality rates in the solvent control group and in treatment Groups 5 and 6 are

$N_2 = N_5 = 59.2, N_6 = 5.8$, $\hat{p}_2 = 2.2/59.2 = 0.037$, $\hat{p}_5 = 5.2/59.2 = 0.088$, $\hat{p}_6 = 0.50$.

Substituting $\hat{p}_2, \hat{q}_2, \hat{p}_5, \hat{q}_5, \hat{p}_6, \hat{q}_6$ for the corresponding parameters in the confidence interval expression we obtain the following results:

Group 5 vs Solvent Control: $\hat{\phi} = \ell n(\hat{p}_5/\hat{p}_2) = \ell n 2.364 = 0.860$

95% confidence interval on $\phi = (0.860 - 1.96(0.784),\ 0.860 + 1.96(0.784)) = (-0.677, 2.397)$

95% confidence interval on $p_5/p_2 = (e^{-0.677},\ e^{2.397}) = (0.51, 10.99)$.

Group 6 vs Solvent Control: $\hat{\phi} = \ell n\, 13.514 = 2.604$

95% confidence interval on $\hat{\phi} = (2.604 - 1.96(0.589), 2.604 + 1.96(0.589)) = (1.448, 3.759)$

95% confidence interval on $p_6/p_2 = (e^{1.448}, e^{3.759}) = (4.26, 42.91)$

We conclude from these confidence interval calculations that:

- p_6 is significantly different from p_2 at the 5% level of significance but p_5 is not (since the latter confidence interval includes 1 while the former does not).
- p_6 is substantially greater than p_2 (since the lower confidence bound on the ratio is 4.26).
- Neither p_5/p_2 nor p_6/p_2 are determined very precisely. The data are compatible with p_5 being just half of p_2 or being ten times greater than p_2. The data are compatible with p_6 being anywhere between 4.3 and 43 times greater than p_2. Thus these ratios are not even determined to an order of magnitude.

It is interesting to note that these results are compatible with those obtained in Section IX, based on various hypothesis testing procedures, where it was shown that Group 5 is the MATC.

With Goulden's Isophorone data there was no adjustment of sample sizes due to beaker-to-beaker heterogeneity. The sample sizes and observed mortality rates among the multiply housed daphnids in the control group and in treatment groups 4, 5, and 6 are $N_1 = N_4 = N_5 = N_6 = 15$, $\hat{p}_1 = 0.133, \hat{p}_4 = 0.200, \hat{p}_5 = 0.533, \hat{p}_6 = 0.867$. Thus the point estimates of and asymptotic 95% confidence intervals on $p_4/p_1, p_5/p_1, p_6/p_1$ are:

Group 4 vs Control: $\hat{p}_4/\hat{p}_1 = 1.50$
95% confidence interval on $p_4/p_1 = (0.29, 7.76)$

Group 5 vs Control: $\hat{p}_5/\hat{p}_1 = 4.00$
95% confidence interval on $p_5/p_1 = (1.01, 15.87)$

Group 6 vs Control: $\hat{p}_6/\hat{p}_1 = 6.50$
95% confidence interval on $p_6/p_1 = (1.76, 24.09)$

We conclude from these confidence interval calculations that:

- p_6 and p_5 are significantly different from p_1 at the 5% level of significance but p_4 is not.
- p_6 and p_5 may be about the same as p_1 or may be an order of magnitude greater.
- None of the ratios $p_4/p_1, p_5/p_1$, or p_6/p_1 are determined very precisely. These ratios cannot be determined by these data even to an order of magnitude.

Poisson Approach

The previous paragraph discussed the construction of confidence intervals based on asymptotic normal theory. The normal approximation to the binomial distribution is certainly valid if $Np>5$ and $Nq>5$ for each group. For several of the groups however, the expected frequencies were as low as 2. Thus the validity of the asymptotic theory is brought into question.

Another approximate approach to constructing confidence intervals on ratios of mortality probabilities is based on the Poisson approximation to the binomial distribution. This approximation is best when both the control group and treatment group probabilities are small, about 0.10 or less. The approximation holds for both small and large sample sizes. Feder and Collins [1], Subsection XIIID, discuss this approach. Following Nelson [31] they state that if X_c, N_c and X_t, N_t are the (adjusted) responses and sample sizes in the control and treatment groups respectively, if p_c, p_t are the response probabilities in these groups, and if $\lambda_c = N_c p_c, \lambda_t = N_t p_t$, then an approximate $1-\alpha$ level two sided confidence interval on λ_t/λ_c is

$$\left[\frac{X_t}{X_c+1}\ \frac{1}{F(2X_c+2,2X_t;1-\alpha_1)},\ \frac{X_t+1}{X_c}\ F(2X_t+2,2X_c;1-\alpha_2)\right]$$

where $F(\nu_1,\nu_2;\gamma)$ represents the upper γ-th percentile of the F-distribution with degrees of freedom ν_1,ν_2 and where $\alpha_1+\alpha_2=\alpha$. Now

$$\lambda_t/\lambda_c=(N_t p_t)/(N_c p_c)=(N_t/N_c)(p_t/p_c)$$

Thus multiplying the above confidence bounds by N_c/N_t yields confidence bounds on p_t/p_c. Note that if $X_t=0$ the lower bound is 0 while if $X_c=0$ the upper bound is infinite.

Nelson [31] presents charts which facilitate the construction of two sided 90%, 95%, or 99% confidence intervals. These charts are illustrated and their use is discussed in Feder and Collins.

We illustrate the Poisson based confidence interval procedure with several examples. First consider the comparison of treatment group 5 with the solvent control group in LeBlanc's Test A. In this case $N_2=N_5=59.2$, $X_2=2.2$, $X_5=5.2$. Thus $\hat{p}_2, \hat{p}_5$ are both less than 0.10. As indicated in the discussion on asymptotic confidence intervals,

$$\hat{p}_5/\hat{p}_2=2.364$$

Substituting X_5, X_2 for X_t, X_c respectively and $\alpha_1=\alpha_2=0.025$ in the confidence interval expression, we calculate an approximate 95% confidence interval on p_5/p_2. Namely

$$\left[\frac{5.2}{3.2}\ \frac{1}{F(6.4,10.4;.975)},\ \frac{6.2}{2.2}\ F(12.4,4.4;.975)\right] = \left[\frac{5.2}{3.2}\ \frac{1}{3.946},\ \frac{6.2}{2.2}\ 3.98\right] = (0.41,11.22)$$

The percentiles of the F-distribution with nonintegral degrees of freedom were obtained by linear interpolation.

This confidence interval is qualitatively similar to but slightly wider than the asymptotic theory confidence interval.

We next consider the comparison of treatment Group 4 with the control group in Goulden's Isophorone test. The observed mortality rates in these groups are 0.200 and 0.133 respectively so we are stretching the Poisson theory a bit. However, the errors made should be on the conservative side (i.e., overly long intervals). In this example $N_1=N_4=15, X_1=2, X_4=3$. Thus

$$\hat{p}_4/\hat{p}_1=1.50$$

Substituting X_4, X_1 for X_t, X_c and $\alpha_1=\alpha_2=0.025$ in the confidence internal expression, we obtain

$$\left[\frac{3}{3}\frac{1}{F(6,6;.975)}, \frac{4}{2}F(8,4;0.975)\right]=\left[\frac{1}{5.82}, 2(8.94)\right]=(0.17,17.84) \quad .$$

This interval is very much wider than the asymptotic confidence interval. Thus the Poisson interval is not consistent with the asymptotic interval in this example. We will be able to determine the relative merits of the asymptotic and Poisson confidence intervals for this example when we consider below the calculation of exact confidence intervals.

Exact, Small Sample, Conditional Approach

If the sample sizes are not sufficiently large to apply the asymptotic confidence interval procedure and if response proportions are not sufficiently small to apply the Poisson confidence interval procedure, then confidence interval comparisons between the treatment groups and control group(s) can be made by an exact, small sample procedure based on the nonull distribution of Fisher's exact test, conditional on the margins of a 2-by-2 contingency table.

Let p_t, p_c denote the response (mortality) probabilities in the treatment and control groups respectively and let $q=1-p$. Feder and Collins [1], Subsection XIIIC, discuss procedures for placing exact, small sample confidence intervals on the odds ratio

$$\rho\equiv\frac{p_c/q_c}{p_t/q_t}$$

based on an algorithm by Thomas [32]. Thomas' algorithm has been implemented in EXAX2 [2]. Thomas and Gart [30] have extended this procedure to place exact confidence intervals on differences and ratios of response probabilities, based on the confidence interval on the odds ratio. They present tables of such confidence intervals for a wide variety of possible outcomes of 2-by-2 contingency tables.

We illustrate the small sample, conditional approach with several examples based on Goulden's Isophorone test. The odds ratios and associated 95% confidence intervals are calculated by EXAX2. Results for pairwise comparisons of Groups 4, 5, and 6 with the control group are shown in Figure XIII.1.

Group 4 vs Control: $\hat{\rho}=0.615$

95% confidence interval on ρ is $(0.045,6.484)\equiv(\rho_L,\rho_U)$

Let x_L,x_U denote the numbers of dead <u>control</u> animals that would correspond to odds ratios ρ_L,ρ_U conditional on the marginals of the table. Following Thomas and Gart we have (for $m=5,N_1=N_4=15$)

$$\rho_L \equiv \frac{x_L}{15-x_L}\Big/\frac{5-x_L}{15-(5-x_L)} \qquad \rho_U \equiv \frac{x_U}{15-x_U}\Big/\frac{5-x_U}{15-(5-x_U)}$$

This results in the quadratic equations

$$(\rho_L-1)x_L^2-(10+20\rho_L)x_L+75\rho_L=0$$

$$(\rho_U-1)x_U^2-(10+20\rho_U)x_U+75\rho_U=0$$

Substituting the values of ρ_L,ρ_U, solving the quadratic equations, and retaining the roots that lie between 0 and 5 yields

$$x_L=0.3008$$

$$x_U=4.1614$$

These values yield confidence bounds on p_4/p_1 by noting that p_4/p_1 is estimated by $(5-x)/x$. Thus, $((5-x_U)/x_U,(5-x_L)/x_L)$ constitutes an exact, small sample $1-\alpha$ level two sided confidence interval on p_4/p_1. This yields the interval (0.20,15.62). This interval is a bit shorter than the Poisson based interval, but is very similar. It appears that the asymptotic interval is too short in this example. However the asymptotic, Poisson, and conditional intervals all lead to the same qualitative conclusions that there is no statistical evidence that p_1 and p_4 differ and that the ratio p_4/p_1 cannot be determined very precisely based on just the pairwise treatment-control comparison.

Group 5 vs Control: $\hat{\rho}=0.135$

95% confidence interval on ρ is $(0.0118,\ 1.0012)\equiv(\rho_L,\rho_U)$

Let x_L, x_U have the same interpretations as in the comparison of Group 4 with the control. Using the same procedure as discussed there, we obtain the quadratic equations

$$(\rho_L-1)x_L^2-(5+25\rho_L)x_L+150\rho_L=0$$

$$(\rho_U-1)x_U^2-(5+25\rho_U)x_U+150\rho_U=0$$

This yields x_L=0.3168, x_U=5.0019. Since p_5/p_1 is estimated by (10-x)/x, we obtain the 95% two sided confidence interval on p_5/p_1 to be $(10-x_U)/x_U,(10-x_L)/x_L)$=(0.999,30.57). The lower end point of this interval is in good agreement with the asymptotic interval but the upper endpoint is much greater. However, both intervals lead to the same qualitative conclusions, namely that p_5 is statistically significantly greater than p_1 (at the one sided 0.025 level) and the ratio p_5/p_1 cannot be determined very precisely based on the pairwise treatment-control comparison.

Group 6 vs Control: $\hat{\rho}$=0.024

95% confidence interval on ρ is (0.0018,0.2511) $\equiv (\rho_L,\rho_U)$

Using the same relations discussed previously, x_L and x_U are determined as the roots of the quadratic equations

$$(\rho_L-1)x_L^2-30\rho_L x_L+225\rho_L=0$$
$$(\rho_U-1)x_U^2-30\rho_U x_U+225\rho_U=0$$

that lie between 0 and 15. This yields x_L=0.6063, x_U=5.0076. Since p_6/p_1 is estimated by (15-x)/x, we obtain the 95% two sided confidence interval on p_6/p_1 to be $((15-x_U)/x_U,(15-x_L)/x_L))$=(1.995,23.74). This interval is in rather good agreement with that based on asymptotic theory. This confidence interval shows that there is strong statistical evidence that p_6 is greater than p_1; however the ratio p_6/p_1 cannot be determined very precisely based on the pairwise treatment-control comparison.

We have discussed a procedure for constructing exact, small sample confidence intervals on the ratios of mortality probabilities and have compared the results from this procedure to confidence intervals based on Poisson or on asymptotic theory. The asymptotic confidence intervals appear to be too short when the group sizes are as small as those in the Goulden test (N=15) and the mortality probabilities are relatively small. However, all the confidence interval procedures yielded the same qualitative conclusions. All the confidence intervals constructed indicated that ratios of mortality probabilities could not be determined very precisely based on unadjusted pairwise treatment-control comparisons with the numbers of daphnids per group and the magnitudes of mortality probabilities encountered in these tests. Note that we have not adjusted for simultaneity. Such adjustments can be easily carried out using Bonferroni's method.

The preceding approach can also be easily adapted to constructing exact, small sample conditional confidence intervals on the <u>differences</u> between the treatment group mortality rates and the control group rate. Again, following the notation in Thomas and Gart [30] we let N_t,N_c denote the sample sizes in the treatment and control groups respectively and X_t, X_c denote the number of dead animals in these groups. Conditional on X_t+X_c=m, the difference between the proportions dead in the treatment group and in the control group is estimated by

$$\hat{\Delta} = \frac{m-X_c}{N_t} - \frac{X_c}{N_c}$$

Let x_L, x_U denote the numbers of dead control animals that would correspond to the lower and upper 95% confidence limits ρ_L, ρ_U on the odds ratio. Then

$$\left(\frac{m-x_U}{N_t} - \frac{x_U}{N_c}, \frac{m-x_L}{N_t} - \frac{x_L}{N_c}\right)$$

constitutes an exact, small sample $1-\alpha$ level two sided confidence interval on p_t-p_c. Substituting the values of N_c, N_t, m, x_L, x_U appropriate for the treatment group-control group comparisons in Goulden's Isophorone Test, we obtain:

Group 4 vs Control($m=5, N_1=N_4=15, x_L=0.3008, x_U=4.1614$): 95% confidence interval on p_4-p_1 is (-0.22,0.29) and the point estimate is 0.067. Thus there is no statistical evidence that p_1 and p_4 differ.

Group 5 vs Control ($m=10, N_1=N_5=15, x_L=0.3168, x_U=5.0019$): 95% confidence interval on p_5-p_1, is (-0.0003,0.62) and the point estimate is 0.40. Thus p_5 is significantly greater than p_1 at the 0.025 level, but we cannot determine the difference very precisely.

Group 6 vs Control ($m=15, N_1=N_6=15, x_L=0.6063, x_U=5.0076$): 95% confidence interval on p_6-p_1 is (0.33,0.92) and the point estimate is 0.73. Thus p_6 is significantly greater than p_1, but the difference cannot be determined very precisely.

C. MORTALITY--SMOOTHED RESPONSES

The confidence intervals in the previous subsection were based on comparisons of the actually observed response rates in the treatment and control groups, without imposing any structure on the mortality probabilities. However a number of assumptions about the behavior of these probabilities with increasing concentration may be quite reasonable. It was seen in the preliminary plots in Section II that the mortality rates generally increase with increasing concentration, they generally increase in a smooth manner, and the trend curves are generally S-shaped. This is typical behavior that is observed for such responses, with many different compounds and many different animal species.

Such structure can be accounted for in the construction of confidence intervals on treatment group-control group differences. One way to do this is to fit regression models to describe the trends in mortality rates and base confidence interval calculations on predictions from these models. One commonly used model is the probit model. Probit models in concentration or in log concentration were fitted to the mortality data in Section X. In this section we use the results from these probit fits to construct confidence intervals on pairwise differences between the treatment group and control group mortality rates. Such confidence intervals would be

expected to be more precise than the intervals constructed in the previous subsections. However they are based on a greater number of assumptions. A detailed comparison of the theoretical properties of the confidence intervals based on the smoothed probability estimates with those based on the unsmoothed estimates might be carried out, but is beyond the scope of this report.

It should be noted that the inferences discussed in this subsection are based on asymptotic theory. Since the predictions of the regression models are based on averaging responses from all the groups in various ways, it might be expected that the asymptotic theory is more valid for the smoothed estimates than for the unsmoothed estimates. This too needs to be investigated in greater detail.

The confidence intervals in this subsection were constructed to compare differences between treatment group and control group mortality rates. Directly analogous procedures can be used to construct confidence intervals on the ratios of these rates.

The confidence intervals are based on the standard three-parameter probit models (in either logarithmic or untransformed concentration levels) that were fitted to the data in Section X. These models can be expressed as

$$p(\text{conc})=p_o+(1-p_o)\Phi(\beta_o+\beta_1(z-m))$$

where p_o,p(conc) are the background mortality rate and the rate at concentration conc respectively, z is either concentration or $\log_{10}$(concentration), m is a fixed centering constant, $\Phi(\cdot)$ is the standard normal c.d.f. and p_o,β_o,β_1 are unknown parameters to be estimated from the model fit to the data. The confidence interval calculations are based on the estimated parameters from this model and their estimated variances and covariances. The details of the procedure are contained in Appendix AXIII. We apply the procedure below to the 21-day mortality responses from several of the data sets.

LeBlanc Test A--Solvent Control Group--Logarithmic Concentration

Here $z=\log_{10}(\text{conc})$, m=-2.113. From Figure X.1 we obtain

$\hat{p}_o=0.08313$ $\hat{\beta}_o=2.52089$ $\hat{\beta}_1=8.27337$

$\hat{\sigma}(\hat{p}_o)=0.018428$ $\hat{\sigma}(\hat{\beta}_o)=1.20098$ $\hat{\sigma}(\hat{\beta}_1)=2.49065$

$$\hat{R}=\begin{pmatrix} 1.0000 & -0.2071 & 0.1886 \\ -0.2071 & 1.0000 & -0.9546 \\ 0.1886 & -0.9546 & 1.0000 \end{pmatrix}$$

The control group is Group 2. The differences and their estimated standard errors are

Group	3	4	5	6	7
Difference	0.0000*	0.0000*	0.004	0.419	0.905
Std Error	0.0000*	0.0000*	0.0131	0.2025	0.0224

Thus 95% confidence intervals on the treatment group-control group differences are

Group 3 vs Control (0.0000,0.0000)
Group 4 vs Control (0.0000,0.0000)
Group 5 vs Control (-0.021,0.030)
Group 6 vs Control (0.022,0.816)
Group 7 vs Control (0.861,0.949)

There is thus statistical evidence that Groups 6 and 7 have greater mortality rates than the control group. However the extent of these differences cannot be well determined. Note that the Group 6 vs Control confidence interval on p_6-p_2 above gives a somewhat different impression of the relation between these probabilities than the asymptotic confidence interval on p_6/p_2 in Subsection B. The interval in this subsection suggests less information about the extent of difference between p_6 and p_2 than does the interval in Subsection B. Most of the discrepancy is due to the fact that the smoothed background mortality estimate is 0.083 while the unsmoothed estimate is 0.038. The remainder of the discrepancy is probably due to slightly different assumptions built into the asymptotic distributions that were used. While both intervals lead to the same qualitative differences, the implications of the interpretational differences of the lower endpoints need to be further investigated. Subjectively, I would prefer the interval in this subsection since it uses more information from the data.

LeBlanc Test B--Combined Control Groups--Untransformed Concentration

Here z=conc, m=0.0250. From Figure X.8 we obtain

$\hat{p}_o$=0.10944 $\hat{\beta}_o$=-3.37287 $\hat{\beta}_1$=15.9695

$\hat{\sigma}(\hat{p}_o)$=0.020227 $\hat{\sigma}(\hat{\beta}_o)$=2.80322 $\hat{\sigma}(\hat{\beta}_1)$=12.9349

$$\hat{R}=\begin{pmatrix} 1.0000 & -0.5474 & 0.5921 \\ -0.5474 & 1.0000 & -0.9868 \\ 0.5291 & -0.9868 & 1.0000 \end{pmatrix}$$

The slope and intercept are not well determined in this example. Their estimates are highly intercorrelated. This is undoubtedly due to the nearly constant dose response relation, except for Group 7.

---

* These values are zero to at least four decimal places.

The control group is Groups 1 and 2 combined. The differences and their estimated standard errors are

Group	3	4	5	6	7
Difference	0.0001	0.0003	0.0017	0.0096	0.5022
Std Error	0.0011	0.0031	0.0124	0.0489	0.1637

Except for Group 7, the treatment group-control group differences are negligible. 95% confidence intervals on the treatment group control group differences are

Group 3 vs Control (-0.002,0.002)
Group 4 vs Control (-0.006,0.006)
Group 5 vs Control (-0.023,0.026)
Group 6 vs Control (-0.086,0.105)
Group 7 vs Control (0.181,0.823)

There is thus statistical evidence that Group 7 has greater mortality rates then the control group. However, the extent of this difference cannot be well determined.

Goulden Isophorone--Logarithmic Concentration

Here $z=\log_{10}(\text{conc})$, $m=1.8352$. From Figure X.11 we obtain

$\hat{p}_o=0.09058$ $\hat{\beta}_o=-2.50889$ $\hat{\beta}_1=7.50132$

$\hat{\sigma}(\hat{p}_o)=0.04258$ $\hat{\sigma}(\hat{\beta}_o)=0.95596$ $\hat{\sigma}(\hat{\beta}_1)=2.51776$

$$\hat{R}=\begin{pmatrix} 1.0000 & -0.3238 & 0.2590 \\ -0.3238 & 1.0000 & -0.9633 \\ 0.2590 & -0.9633 & 1.0000 \end{pmatrix}$$

The control group is Group 1. The differences and their estimated standard errors are

Group	2	3	4	5	6
Difference	0.0000	0.0002	0.0924	0.4722	0.7619
Std Error	0.0000	0.0009	0.0932	0.1051	0.0884

95% confidence intervals on the treatment group-control group differences are

Group 2 vs Control (negligible)
Group 3 vs Control (-0.002,0.002)
Group 4 vs Control (-0.090,0.275)
Group 5 vs Control (0.266,0.678)
Group 6 vs Control (0.589,0.935)

There is statistical evidence that Groups 5 and 6 have greater mortality rates than the control group. However the extent of these differences cannot be well determined.

It is interesting to compare these intervals with those obtained in Subsection B based on exact, small sample theory using the unsmoothed mortality estimates. Direct comparisons are available for Groups 4, 5, and 6 in Goulden's Isophorone Test.

	Exact, Small Sample Confidence Intervals, Unsmoothed Mortality Estimates	Asymptotic Confidence Interval, Probit Based Mortality Estimates
Group 4 vs Control	(-0.22,0.29),0.067	(-0.090,0.275),0.092
Group 5 vs Control	(-0.0003,0.62),0.40	(0.266,0.678),0.47
Group 6 vs Control	(0.33,0.92),0.73	(0.589,0.935),0.76

The smoothed and unsmoothed point estimates are similar. However the asymptotic confidence intervals are much shorter than the small sample confidence intervals. The discrepancies are particularly at the lower endpoints of these intervals. The reason for good agreement of the upper confidence bounds but poor agreement of the lower confidence bounds is not well understood and should be studied further.

D. LENGTH--UNSMOOTHED RESPONSES

Subsections B and C were concerned with various approaches to constructing confidence intervals for the comparisons of treatment group and control group mortality rates. Confidence intervals can also be constructed to compare length and reproduction responses between treatment and control groups. Since the procedures for length and for reproduction are quite similar, we consider lengths only. In this subsection we work with unsmoothed lengths and in the next subsection we work with lengths smoothed by regression models. Recall that in several data sets we deleted groups from the length and reproduction comparisons because of excessive mortality. This includes Group 7 in LeBlanc's Test A and Group 6 in Goulden's Isophorone test.

Nonsimultaneous confidence intervals are based on the t-distribution. Adjustments for simultaneity can be based on Dunnett's procedure in the balanced case (i.e., equal numbers of beakers per group and equal numbers of daphnids per beaker) or more generally on Bonferroni's procedure. We will not pursue either procedure here. We consider several examples.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted

We use the average lengths and estimated standard errors contained in Subsection XI.C. The standard error estimates are based on the interaction mean square in the analysis of variance fit discussed in Subsections IV-A and V-D. The average lengths and estimated standard errors are (see Subsection V-D for details)

	Control				
Group	2	3	4	5	6
Average	5.2753	4.9743	5.1027	4.9658	5.0000
Std Error	0.059	0.062	0.061	0.061	0.080

We associate 98 degrees of freedom with the interaction mean square (after augmenting information from the within beaker mean square).

The 1-α level two sided nonsimultaneous confidence intervals are $\bar{Y}_i - \bar{Y}_o \pm t_o(\hat{\sigma}_i^2+\hat{\sigma}_o^2)^{1/2}$, where $\bar{Y}_i, \bar{Y}_o$ are the average lengths, where t_o is the upper $\alpha/2$ point of the t-distribution with 98 degrees of freedom, and $\hat{\sigma}_i, \hat{\sigma}_o$ are the estimated standard errors of the treatment and control group averages respectively. If 1-α=0.95, the t factor is 1.99. Thus the confidence intervals are

Group 3 vs Control (-0.471,-0.131)
Group 4 vs Control (-0.341,-0.004)
Group 5 vs Control (-0.478,-0.141)
Group 6 vs Control (-0.473,-0.077)

Since none of these intervals contain 0, they provide statistical evidence that the average 21-day length in the solvent control group is significantly greater than those in the treatment groups. The average differences are between a tenth and a half of a millimeter. Whether or not such reductions in lengths are of biological significance is a separate issue. This result is in direct agreement with the appearance of Figure II.26 and with the multiple comparisons calculations carried out in Subsection XI.C.

Chapman--Beryllium--Combined Control Groups

Analyses are carried out only on lengths corresponding to daphnids that survived to the end of the test. Since the test was carried out with just one daphnid per beaker, there is no complication due to beaker-to-beaker variation within groups. This component of variation is confounded with the daphnid-to-daphnid variation.

The estimated standard deviation, based on the residual sum of squares from a one-way analysis of variance (not shown) is $\hat{\sigma}$=0.325 with 56 degrees of freedom. The average lengths per group, numbers of daphnids, and standard errors are

	Control						
Group	1 and 2	3	4	5	6	7	8
Number	18	8	6	6	9	10	6
Average	4.278	4.125	4.000	3.817	3.900	3.590	3.650
Std Error	0.077	0.115	0.133	0.133	0.108	0.103	0.133

Nonsimultaneous 95% confidence intervals are calculated just like in the LeBlanc Test A example. These confidence intervals are

Group 3 vs Control	(-0.429,0.124)
Group 4 vs Control	(-0.585,0.029)
Group 5 vs Control	(-0.768,-0.154)
Group 6 vs Control	(-0.643,-0.113)
Group 7 vs Control	(-0.945,-0.431)
Group 8 vs Control	(-0.935,-0.321)

These intervals provide statistical evidence that the average lengths in Groups 5-8 are statistically significantly lower than those in the control group. This is in reasonable agreement with the appearance of Figure II.32.

E. LENGTH--SMOOTHED RESPONSES

The confidence intervals in the previous subsection were based on comparisons of the actually observed average lengths within each group, without imposing any structure on these averages. Regression models were fitted to the lengths in Subsection XII.C that assumed smooth trends in average lengths with increasing concentrations. It is interesting to note that these trends are not necessarily monotone. That is, low toxicant concentrations sometimes enhance average lengths.

The confidence intervals in this subsection were constructed to compare differences between treatment group and control group average lengths. They are based on polynomial regression models (linear or quadratic) in $\log_{10}$(concentration). Since the control group is generally far removed from the treatment groups in terms of $\log_{10}$(concentration), it was felt advisable to fit the polynomials only to the treatment group responses. The control group responses are unadjusted. The specific form of the models fitted is discussed in detail in Subsection XII.A. See that discussion for further details. We utilize the results of these models in the calculations below.

We illustrate the construction of confidence intervals on treatment group-control group average differences with several examples.

LeBlanc Test A--Solvent Control Group--Group 7 Deleted

There are four treatment groups and a control group. A quadratic regression model in $\log_{10}$(concentration) was fitted to the treatment group responses. The results are shown in Figure XII.17. The quadratic term in Figure XII.17 has the wrong sign (positive), is small, and is nonsignificant. The plot of length vs group in Figure II.26 shows no trend among the treatment groups. The quadratic term was deleted from the model and a straight line trend was fitted to the treatment groups. The results are shown in Figure XII.19. The slope coefficient is essentially zero and is nonsignificant. Based on this and the appearance of Figure II.26, we conclude that there is no trend in average lengths among the four treatment

groups. We thus calculate an overall average length across the treatment groups for comparison with that in the control group. The average length among the 256 surviving daphnids in the 4 treatment groups is 5.0125 mm with a standard error (based on the overall interaction mean square) of 0.0326. (The average of the 16 beaker averages is 5.0231 with a standard error of .035, based on these 16 values.) The average length among the 77 surviving daphnids in the solvent control group is 5.2753 mm with a standard error (based on the overall interaction mean square) of 0.059. (The average of the four beaker averages is 5.2775 mm with a standard error of 0.029, based on those 4 values. The averages are similar but the standard errors differ. We use the larger value below, but this should be considered further.)

A 95% confidence interval on the difference of the treatment group and solvent control group average lengths is

$$(5.0125-5.2753)\pm 2(.0326^2+.059^2)^{1/2}=(-0.40,\ -0.13)$$

This interval is tighter than the individual confidence intervals calculated in the previous subsection and it summarizes the trend information more succinctly. We feel that this is a better representation of the results.

Chapman--Beryllium--Combined Control Groups

There are six treatment groups and a (combined) control group. Analyses are carried out only on lengths corresponding to daphnids that survived to the end of the test. A quadratic trend model in $\log_{10}$(concentration) was fitted to the treatment group responses. The results of the fit are shown in Figure XII.21. It is evident from Figures XII.21 and XII.22 that the quadratic trend is negligible. The quadratic term was thus deleted from the model and a straight line model was fitted to the data. The results of this fit are shown in Figure XII.23. The linear term is highly significant and we use this model to construct confidence intervals.

The parameterization of the regression model is discussed in detail in Section XII. The portion, $\beta_0+\beta_1 x$, of the model represents the difference between the predicted average length at x ($=\log_{10}$(concentration)) and the average length in the control group. This is estimated for the i-th treatment group as $\hat{\beta}_0+\hat{\beta}_1 x_i$ where $\hat{\beta}_0, \hat{\beta}_1$ are the I and IX coefficients in Figure XII.23 and x_i is 0.4150, 0.7931, 1.1377, 1.4294, 1.7370, 2.0456 for treatment Groups 3, 4, 5, 6, 7, 8 respectively. The standard error of $\hat{\beta}_0+\hat{\beta}_1 x_i$ is estimated as

$$\hat{\sigma}_i=[(1,x_i)\hat{\Sigma}(1,x_i)']^{1/2}$$

where $\hat{\Sigma}$ is the estimated variance-covariance matrix given at the bottom of Figure XII.23. A 1-α two sided nonsimultaneous confidence interval on the treatment-control difference is $\hat{\beta}_0+\hat{\beta}_1 x_i \pm t\hat{\sigma}_i$, where t is the upper $\alpha/2$ percentile of the distribution with the appropriate number of degrees of freedom. In our case α=0.05, d.f.=56, t=2.00. The confidence intervals on the differences for the individual treatment groups are:

Group 3 vs Control	(-0.391,0.078)
Group 4 vs Control	(-0.478,-0.082)
Group 5 vs Control	(-0.573,-0.212)
Group 6 vs Control	(-0.699,-0.307)
Group 7 vs Control	(-0.785,-0.392)
Group 8 vs Control	(-0.913,-0.465)

These intervals are qualitatively similar to the unsmoothed intervals calculated in the previous subsection. However they are somewhat shorter, since they incorporate information from all the groups. Groups 5-8 are statistically significantly lower than the control group. Group 4 is seen to be borderline significant on the basis of either the smoothed or unsmoothed intervals.

XIV. EXPERIMENTAL DESIGN CONSIDERATIONS

In this section we consider a number of aspects of the design of chronic Daphnia toxicity tests. Topics discussed include limitations on test size, precision to be expected as a function of test size, allocation of daphnids among beakers and beakers among groups, criteria for judging the adequacy of a test, and blocking considerations. Feder and Collins [1], Section XIII discuss some experimental design issues related to early life stage tests with Fathead Minnows. Portions of the discussion in this section extend that material and adapt it to Daphnia tests.

A. INTRODUCTION

The principal limitations on the size of a toxicity test are the amount of dilution water that can be supplied by the particular proportional diluter being used, the number of replicate beakers per group into which this water can practicably be split in a uniform manner, and the maximum number of daphnids that can be accommodated in the available amount of water and that can be monitored for survival, reproduction, and length with the available amount of laboratory personnel and technology. The extent of these limitations depends on the available equipment and people resources and this will vary considerably among test facilities. Thus no absolutes can be stated about universal numbers of daphnids, numbers of beakers, etc., that need be used. However, we can estimate the precision to be expected as a function of test size and we can state some general guidelines for allocating daphnids among beakers and allocating beakers among groups.

B. ALLOCATION OF DAPHNIDS AMONG BEAKERS WITHIN GROUPS

Experimental equipment, facilities, and technique place upper bounds on the numbers of beakers that can be used in each treatment or control group. Water supply and technician availability place upper limits on the number of Daphnia that can be used. We recommend that as many daphnids as feasible be used for studying toxicant effects on survival. Furthermore as many beakers as possible should be used within each test group, so that the numbers of daphnids per beaker can be made as small as possible. However equal numbers of daphnids should be allocated to each beaker within each test group, so as to keep effects of competition, food and oxygen supply, contagion, and handling as constant as possible across beakers.

Utilizing larger numbers of replicate beakers per group with smaller numbers of daphnids per beaker has many advantages. Some of these are:

1. There will be less competition among the daphnids within each beaker. This should result in healthier daphnids, should reduce nontoxicant related mortality and morbidity, and should thereby

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 4

ALPHAL = .025 ALPHAU = .025

		LIVE	DEAD	TOTAL
TRT	1	13.000	2.000	15.000
TRT	4	12.000	3.000	15.000

TOTAL LIVE = 25.000 , TOTAL DEAD = 5.000 , GRAND TOTAL = 30.000

ESTIMATE OF (P1Q2)/(P2Q1) = .6153846154
95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (.0448646545, 6.4838980955)
PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = .50000

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 5

ALPHAL = .025 ALPHAU = .025

		LIVE	DEAD	TOTAL
TRT	1	13.000	2.000	15.000
TRT	5	7.000	8.000	15.000

TOTAL LIVE = 20.000 , TOTAL DEAD = 10.000 , GRAND TOTAL = 30.000

ESTIMATE OF (P1Q2)/(P2Q1) = .1346153846
95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (.0118455887, 1.0011610144)
PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = .02509

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 6

ALPHAL = .025 ALPHAU = .025

		LIVE	DEAD	TOTAL
TRT	1	13.000	2.000	15.000
TRT	6	2.000	13.000	15.000

TOTAL LIVE = 15.000 , TOTAL DEAD = 15.000 , GRAND TOTAL = 30.000

ESTIMATE OF (P1Q2)/(P2Q1) = .0236686391
95.00 % ACA-SIMULTANEOUS CONFIDENCE INTERVAL = (.0017741919, .2511444892)
PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = .00007

NOTE: THIS IS A ONE-TAILED PROBABILITY. IF CONTROL MORTALITY EXCEEDS TREATMENT MORTALITY, THEN 1 MINUS SIGNIFICANCE LEVEL IS APPROPRIATE.

Figure XIII.1. Estimated odds ratios and associated 95% confidence intervals for pairwise comparisons of mortality rates in groups 4, 5 and 6 with the control. Goulden isophorone.

improve sensitivity for determining toxicant related effects on survival, growth, and reproduction.

2. There will be less contagion among the daphnids within each beaker. Thus fungii, bacteria, etc., that can invade a beaker will affect fewer daphnids. Similarly any fluctuations in experimental conditions in the beaker will affect fewer daphnids.

3. Estimates of effects within groups are based on averages of the observed average effects within each beaker. The presence of beaker-to-beaker heterogeneity degrades the precision of such averages. However, averaging over replicate beakers tend to balance out the beaker-to-beaker heterogeneity. Another way of saying this is that for a given number of daphnids on test per group and for a given degree of beaker-to-beaker heterogeneity per group, the more beakers and the fewer daphnids per beaker that are used, the more precise will be the estimates. In the terminology used in previous sections, the effective sample size per group increases as the number of beakers increase.

4. The number of degrees of freedom for estimating the extent of beaker-to-beaker variation within groups increases with the number of beakers. Thus not only is the effective sample size and therefore the precision increased, but the ability to estimate that precision is improved. Feder and Collins [1], Section XVIII recommend that at least 12 degrees of freedom be available for estimating variability.

The principal limitations on the numbers of beakers that can be used per group are the ability to deliver uniform water quantity and quality to large numbers of beakers and to maintain the beakers under uniform laboratory test conditions. These problems increase as the size of the test increases.

We now perform calculations that illustrate how effective sample sizes per group depend on the allocation of beakers and daphnids within each group. We illustrate such calculations for survival, length, and reproduction responses.

Suppose that a particular test (or control) group contains J beakers and n daphnids per beaker. The total sample size is $N \equiv Jn$. We calculate effective sample sizes as J (and therefore n) varies with N fixed. Effective sample size calculations for determing survival rates are based on Williams' beta binominal model [3] while effective sample sizes for determining average lengths or reproduction are based on components of variance calculations.

We first consider survival. Following the discussion in Feder and Collins [1], Subsection XVIIIB we let X_{ij} denote the number of dead daphnids (e.g., after 21 days) in beaker j of group i. For notational convenience we suppress the subscript i in the discussion below. Thus X_{ij} is referred to as X_j. We assume X_j is binominally distributed with parameters (n, p_j) and p_j in turn has a beta distribution with parameters (α, β). This model allows for random variation of mortality rates among the beakers within

TABLE XIV.1 EFFECTIVE SAMPLE SIZE PER GROUP AS A FUNCTION OF NUMBER OF DAPHNIDS (N) NUMBER OF BEAKERS (J), AND DEGREE OF BEAKER-TO-BEAKER HETEROGENEITY (θ)

N=10	θ	0	0.01	0.02	0.04	0.07	0.10	0.20	0.30	0.50	0.75	1.00	1.50	2.00
J=1		10	9.2	8.5	7.4	6.3	5.5	4.0	3.3	2.5	2.1	1.8	1.6	1.4
2		10	9.6	9.3	8.7	7.9	7.3	6.0	5.2	4.3	3.7	3.3	2.9	2.7
5		10	9.9	9.8	9.6	9.4	9.2	8.6	8.1	7.5	7.0	5.2	6.2	6.0
10		10	10	10	10	10	10	10	10	10	10	10	10	10
N=15														
J=1		15	13.2	11.8	9.8	7.8	6.6	4.5	3.6	2.7	2.1	1.9	1.6	1.5
3		15	14.4	13.9	13.0	11.9	11.0	9.0	7.8	6.4	5.5	5.0	4.4	4.1
5		15	14.7	14.4	13.9	13.2	12.7	11.3	10.3	9.0	8.1	7.5	6.8	6.4
15		15	15	15	15	15	15	15	15	15	15	15	15	15
N=20														
J=1		20	16.8	14.6	11.6	8.9	7.3	4.8	3.7	2.7	2.2	1.9	1.6	1.5
2		20	18.4	17.0	14.9	12.6	11.0	8.0	6.5	5.0	4.1	3.6	3.1	2.9
4		20	19.2	18.6	17.3	15.9	14.7	12.0	10.4	8.6	7.4	6.7	5.9	5.5
10		20	19.8	19.6	19.3	18.8	18.3	17.1	16.3	15.0	14.0	13.3	12.5	12.0
N=30														
J=1		30	23.3	19.1	14.2	10.4	8.3	5.1	3.9	2.8	2.2	1.9	1.6	1.5
2		30	26.4	23.5	19.5	15.7	13.2	9.0	7.1	5.3	4.3	3.8	3.2	2.9
3		30	27.6	25.5	22.3	18.9	16.5	12.0	9.8	7.5	6.2	5.5	4.7	4.3
5		30	28.6	27.3	25.2	22.6	20.6	16.4	16.4	13.9	11.3	9.6	8.6	6.9
10		30	29.4	28.9	27.9	26.5	25.4	22.5	20.5	18.0	16.2	15.0	13.6	12.9
N=40														
J=2		40	33.7	29.1	23.1	17.8	14.7	9.6	7.4	5.5	4.4	3.8	3.2	2.9
4		40	36.7	34.0	29.7	25.2	22.0	16.0	13.0	10.0	8.3	7.3	6.3	5.7
8		40	38.5	37.1	34.7	31.7	29.3	24.0	20.8	17.1	14.7	13.3	11.8	10.9
10		40	38.9	37.8	35.9	33.4	31.4	26.7	23.6	20.0	17.5	16.0	14.3	13.3
20		40	39.6	39.2	38.5	37.5	36.7	34.3	32.5	30.0	28.0	26.7	25.0	24.0
N=50														
J=2		50	40.4	34.0	26.0	19.5	15.7	10.0	7.7	5.6	4.4	3.9	3.3	2.9
5		50	45.9	42.5	37.1	31.5	27.5	20.0	16.3	12.5	10.3	9.1	7.8	7.1
10		50	48.1	46.4	43.3	39.6	36.7	30.0	26.0	21.4	18.4	16.7	14.7	13.6
25		50	49.5	49.0	48.2	46.9	45.8	42.9	40.6	37.5	35.0	33.3	31.3	30.0
N=75														
J=3		75	60.5	51.0	39.0	29.2	23.6	15.0	11.5	8.3	6.7	5.8	4.9	4.4
5		75	65.9	58.9	48.8	39.2	33.0	22.5	17.7	13.2	10.7	9.4	8.0	7.3
15		75	72.1	69.6	65.0	59.4	55.0	45.0	39.0	32.1	27.6	25.0	22.1	20.5
25		75	73.5	72.2	69.6	66.3	63.5	56.3	51.3	45.0	40.4	37.5	34.1	32.1
N=100														
J=2		100	67.3	51.0	34.7	23.8	18.3	10.9	8.1	5.8	4.6	3.9	3.3	3.0
4		100	80.8	68.0	52.0	38.9	31.4	20.0	15.3	11.1	8.9	7.7	6.5	5.9
10		100	91.8	85.0	74.3	62.9	55.0	40.0	32.5	25.0	20.6	18.2	15.6	14.3
20		100	96.2	92.7	86.7	79.3	73.3	60.0	52.0	42.9	38.8	33.3	29.4	27.3
25		100	97.1	94.4	89.7	83.6	78.6	66.7	59.1	50.0	43.8	40.0	35.7	33.3

each group and thereby quantifies the extent of beaker-to-beaker heterogeneity.

Let $\mu \equiv \alpha/(\alpha+\beta)$ $\theta \equiv 1/(\alpha+\beta)$ $0 \leq \mu \leq 1, 0 \leq \theta < \infty$

Then

$$E(p_j) \equiv \mu \qquad Var(p_j) = \mu(1-\mu)\theta/(1+\theta)$$

The unconditional distribution of X_j, accounting for the random variation of p_j, is beta binominal with mean and variance

$$E(X_j) = n\mu \qquad Var(X_j) = n\mu(1-\mu)\frac{1+n\theta}{1+\theta}$$

Thus

$$p = \Sigma_j X_j/N \qquad E(\hat{p}) = \mu \qquad Var(\hat{p}) = \frac{\mu(1-\mu)}{N}\ \frac{1+n\theta}{1+\theta}$$

In Subsection VB we defined the variance inflation factor as

$$K \equiv \frac{Var(\hat{p})}{[\mu(1-\mu)/N]}$$

and the effective sample size as

$$N_{eff} \equiv N/K$$

Thus, under the assumptions of the beta binomial model,

$$K = \frac{1+n\theta}{1+\theta} \qquad \text{and} \quad N_{eff} = \frac{1+\theta}{1+n\theta}$$

The parameter θ quantifies the extent of beaker-to-beaker heterogeneity within groups. As θ approaches 0, there is less and less beaker-to-beaker heterogeneity and so N_{eff} approaches N, the number of daphnids. As θ approaches infinity, there is more and more beaker-to-beaker heterogeneity and so N_{eff} approaches J, the number of beakers. The general situation is a compromise between these two extremes. We also see that for fixed N and fixed θ, N_{eff} increases as n decreases (i.e., as J increases). We tabulate values of N_{eff} in Table XIV.1, corresponding to various combinations of N, J, and θ. Table XIV.1 shows that depending on the specific combinations of N, J, θ the effective sample size can be anywhere between J and N. This has important implications on the sensitivity of the test. The added precision due to having additional beakers is most pronounced when the extent of beaker-to-beaker heterogeneity is relatively large and the number of beakers per group is relatively small.

What values of θ arise on daphnia toxicity tests? A partial answer to this question can be obtained by reference to several of the data sets analyzed in previous sections. No evidence of beaker-to-beaker

heterogeneity was present in either Adams' selenium test or in Goulden's isophorone test. However, the relatively small group sizes in these tests (J=3, n=5) did not permit sensitive determinations of such heterogeneity. LeBlanc's Tests A and B had larger group sizes (J=4, n=20). Both of these tests showed strong statistical evidence of beaker-to-beaker heterogeneity within groups (see Subsection III-B). It was estimated in Subsection V-B that the variance inflation factors are:

LeBlanc Test A	K = 1.35 for Groups 1-5, 7
	K = 13.76 for Group 6
LeBlanc Test B	K = 1.39 for Groups 1-6
	K = 8.91 for Group 7

Using the relation $K = (1+n\theta)/(1+\theta)$ and n=20, we obtain the following estimates of θ.

LeBlanc Test A	θ = 0.02 for Groups 1-5, 7
	θ = 2.04 for Group 6
LeBlanc Test B	θ = 0.02 for Groups 1-6
	θ = 0.71 for Group 7

Referring to the portion of Table XIV.1 corresponding to N=75 and θ=0.02, we see that N_{eff} can vary a moderate amount as J varies between 3 and 25. The effective sample size at J=25 is about 40 percent greater than that at J=3, based of course on the same number of daphnids tested. The situation for θ=0.75 or for θ=2.00 is much more extreme. The effective sample size varies between 6.7 and 40.4 for θ=0.75 and between 4.4 and 32.1 for θ=2.0 as J varies between 3 and 25. Thus having had allocations with more beakers and with fewer daphnids per beaker would have greatly increased the effective sample sizes in Group 6 of Test A and in Group 7 of Test B.

C. SAMPLE SIZE AND POWER CONSIDERATIONS FOR QUALITATIVE SURVIVAL DATA

In the previous subsection we calculated effective group sample sizes as a function of number of beakers per group, number of daphnids per beaker, and extent of beaker-to-beaker heterogeneity. In this subsection we calculate the power to be expected for pairwise comparisons between treatment groups and the control group. These power calculations are based on one sided, two sample hypothesis tests without any smoothing of the treatment group responses by means of fitting regression models. Such smoothing would undoubtedly improve the power of treatment group-control group comparisons.

The power calculations in this subsecton are an extension of those in Feder and Collins [1], Subsection XVIII.B. They are based on adjusting the sample sizes within each group to effective sample sizes and then carrying out comparisons across groups based on per daphnid analyses. The calculations are an extension of those in Feder and Collins, but the sample

sizes are modified to be more applicable to Daphnia tests. Feder and Collins examined the consequences of allocating greater numbers of daphnids to the control group than to each of the treatment groups since the control group enters into each pairwise comparison while each treatment group enters into just one. Thus extra allocation to the control group should improve the resulting power. The calculations in Feder and Collins show that the power is indeed improved; however the extent of improvement is not sufficiently great to warrant the additional logistical and administrative effort involved. Thus the power calculations below are based on the assumption of equal allocations of daphnids in each group (treatment or control). Adjustments for simultaneous testing are ignored in the power calculations.

The power calculations are based on the hypothesis test of $H_o: p=p_o$ vs $H_1: p>p_o$ where p, p_o corresponds to the average mortality rates in the treatment and control groups respectively. We adjust the effective sample sizes down to N_{eff} in each group to account for beaker-to-beaker heterogeneity. We assume that N, J, θ are the same in each group so that N_{eff} is also the same across groups. We estimate p, p_o by the sample mortality rates $\hat{p}, \hat{p}_o$, we carry out the variance stabilizing transformations 2 arc sin $\sqrt{\hat{p}}$ and 2 arc sin $\sqrt{\hat{p}_o}$, and we reject H_o at significance level $\alpha=0.05$ if

$$2 \text{ arc sin } \sqrt{\hat{p}} - 2 \text{ arc sin } \sqrt{\hat{p}_o} > 1.645\sqrt{2/N_{eff}}$$

The power of this test is calculated for various combinatons of N_{eff}, p, p_o that are appropriate for Daphnia tests. Based on the calculations in Table XIV.1, N_{eff} varies between 1 and 100 as N, J, θ vary over different combinations of values. It is assumed that an infinite number of degrees of freedom are available for error estimation.

The observed control group mortality rates in LeBlanc's, Adams', Chapman's, and Goulden's (combined) control groups were 0.09, 0.13, 0.07, 0.10, and 0.13 respectively. We thus assume a range of control group mortality rates between 0.05 and 0.15. The calculations are shown in Table XIV.2.

If we (somewhat arbitrarily) regard 0.80 as a reasonable level of power, we see that tests conducted under ASTM guidelines (i.e., 3 beakers per group to test survival, 5 daphnids per group) are sensitive for distinguishing between 10% and 50% mortality but not for distinguishing between 10% and 30% mortality, even if there is no beaker-to-beaker heterogeneity. Any beaker-to-beaker heterogeneity would degrade the sensitivity of the test since it would reduce the effective sample sizes. For example, for $\theta=0.71$ (the estimated heterogeneity of Group 7 of LeBlanc Test B), Table XIV.1 shows that the effective group size is degraded from 15 to about 6. Table XIV.2 shows that with a group size of 6, the test cannot come close to distinguishing between 10% and 50% mortality rates. Thus even without taking heterogeneity into account, the ASTM guidelines appear to be too minimal for many reasonable survival rate comparisons of biological importance.

TABLE XIV.2 POWER OF ONE SIDED PAIRWISE COMPARISONS OF SURVIVAL RATES BETWEEN CONTROL GROUP AND TREATMENT GROUPS AS A FUNCTION OF N_{eff},p,p_o. α=0.05

N_{eff}	1	2	4	6	10	15	25	50	75	100
p_o=0.05										
p=0.075	0.06	0.06	0.07	0.07	0.08	0.09	0.10	0.13	0.16	0.18
0.10	0.07	0.07	0.09	0.10	0.11	0.13	0.17	0.25	0.32	0.39
0.15	0.08	0.10	0.12	0.15	0.19	0.24	0.34	0.53	0.68	0.79
0.20	0.10	0.12	0.17	0.21	0.28	0.37	0.52	0.77	0.88	0.96
0.30	0.13	0.17	0.26	0.34	0.48	0.62	0.82	0.97	1.00	1.00
0.40	0.16	0.23	0.37	0.48	0.66	0.81	0.95	1.00		
p_o=0.075										
p=0.10	0.06	0.06	0.06	0.07	0.07	0.08	0.09	0.12	0.14	0.15
0.15	0.07	0.08	0.10	0.11	0.13	0.16	0.21	0.33	0.43	0.52
0.20	0.08	0.10	0.13	0.16	0.21	0.27	0.37	0.59	0.74	0.84
0.30	0.11	0.15	0.22	0.28	0.38	0.50	0.69	0.92	0.98	1.00
0.40	0.14	0.20	0.31	0.41	0.57	0.72	0.89	0.99	1.00	
p_o=0.10										
p=0.15	0.06	0.07	0.08	0.08	0.10	0.11	0.13	0.19	0.24	0.28
0.20	0.07	0.09	0.11	0.12	0.16	0.19	0.26	0.41	0.54	0.64
0.30	0.10	0.13	0.18	0.23	0.31	0.41	0.57	0.83	0.94	0.98
0.40	0.13	0.18	0.27	0.35	0.49	0.63	0.82	0.98	1.00	1.00
0.50	0.16	0.24	0.37	0.48	0.67	0.82	0.95	1.00		
p_o=0.15										
p=0.20	0.06	0.07	0.07	0.08	0.09	0.10	0.12	0.16	0.20	0.24
0.30	0.08	0.10	0.13	0.16	0.20	0.26	0.36	0.57	0.72	0.82
0.40	0.11	0.14	0.20	0.26	0.36	0.47	0.65	0.89	0.97	0.99
0.50	0.14	0.29	0.38	0.54	0.68	0.86	0.99	1.00	1.00	

Note that ASTM guidelines suggest running 7 additional beakers per group, with just one daphnid per beaker. These additional daphnids are intended for assessing changes in productivity with increasing concentrations. We recommend that these seven individually housed daphnids per group should not be combined with the 15 multiply housed daphnids per group for assessing mortality rates. The mortality rates will in general differ for singly and for multiply housed daphnids. This was seen to be the case in Goulden's data.

Chapman's beryllium test was conducted with 10 beakers per group and one daphnid per beaker. The intent of this design was to obtain good information about reproduction. Table XIV.2 shows that this test was too small to provide good sensitivity for inferences concerning mortality. (Note that the test was not designed for this purpose.)

LeBlanc's tests were somewhat larger. They consisted of N=80 daphnids per group, divided among J=4 beakers. In most of the groups the degree of beaker-to-beaker heterogeneity was small (i.e., θ=0.02). Table XIV.1 shows that the effective sample size per group is reduced to about N_{eff}=60. Table XIV.2 shows that with effective group sizes of 60, we can expect to distinguish reasonably precisely between 10% and 30% mortality. However in Group 6 of Test A and in Group 7 of Test B the degree of beaker-to-beaker heterogeneity was somewhat greater. The estimated values of θ for these groups are 2.0 and 0.7 respectively. Table XIV.1 shows that with N=80 daphnids and J=4 beakers per group, the effective sample sizes are approximately 6 and 9 respectively. If this degree of beaker-to-beaker heterogeneity had been present in all the groups then Table XIV.2 shows that we could not expect to distinguish well even between mortality rates of 10% and 50%.

We thus see that the sensitivity of a test depends heavily on the extent of beaker-to-beaker heterogeneity within groups as well as on the number of daphnids per group. In turn, the effects of beaker-to-beaker heterogeneity depend on the number of beakers per group. The larger the number of beakers and the smaller the number of daphnids per beaker, the less will be the decrease in precision.

D. UNEQUAL ALLOCATION OF TESTING EFFORT AMONG TREATMENT GROUPS

Feder and Collins [1], Subsection XVIIIE suggested that under certain circumstances it might be sensible to have an asymmetric allocation of beakers and daphnids to the various experimental groups. In particular they state "...if on the basis of either a priori scientific information or previous testing some information was available concerning mortality rates to be expected at the various treatment groups, then unequal allocation of experimental effort would be preferable. In particular at the higher treatment groups, where mortality would be expected to be substantially higher than the control rate, it is easy to detect differences from the control. Thus the experimental effort should be decreased at these groups. At the lower experimental groups, where it is more difficult to detect

differences from the control group, the experimental effort should be increased to improve sensitivity. Thus the degree of experimental effort should in general decrease as the toxicant level increases..."

We discuss here an approach for arriving at an unequal allocation. This procedure leads to equal allocations across groups when there is a priori total ignorance about response levels to be expected. The greater the degree of prior information or beliefs about expected response levels, the more asymmetric will be the allocation.

The use of prior knowledge or information in designing tests seems quite sensible. With all the accumulated experience in the aquatic toxicology literature, there is no need to act as if each test being run is the first test ever. Using equal allocations of beakers within groups is essentially saying that.

We now discuss the details of a procedure for arriving at unequal allocations. We illustrate the procedure with a hypothetical example, but one bearing similarities to several of the data sets we have studied.

Suppose there is a control group and I treatment groups (denoted as Group 0 and Groups 1 to I respectively). Let $p_0, p_1, \ldots, p_I$ denote the average mortality rates in these groups. Based on prior knowledge, information, or belief we can place bounds on these rates. Namely $\ell_0 \le p_0 \le u_0$, $\ell_1 \le p_1 \le u_1$, $\ell_2 \le p_2 \le u_2, \ldots$, $\ell_I \le p_I \le u_I$. The ℓ's and u's are specified quantities. Total ignorance would correspond to $\ell_0 = \ell_1 = \ldots = \ell_I = 0$, $u_0 = u_1 = \ldots = u_I = 1$. We wish to test the hypotheses $H_0: p_i = p_0$ against the one sided alternative $H_1: p_i > p_0$. We stipulate that it is important to reject H_0 whenever p_i is Δ or more above the control group mortality rate (e.g., $\Delta = .10$ or $\Delta = .20$, etc.).

We further assume that cost, experimental, and logistical constraints have placed limits on the total number of beakers and the total number of daphnids to be used in the test as well as the numbers of daphnids to be placed in each beaker. Each beaker throughout the test (that is used for determining survival rates) contains the same number of daphnids, n. We assume that the extent of beaker-to-beaker heterogeneity is constant across groups. Thus n_{eff}, the effective sample size within beakers, is also constant across groups. We allocate the available beakers across groups. In the course of determining the numbers of beakers per group we do not take into account limitations imposed by the proportional diluter apparatus. If a suggested group allocation exceeds the maximum number of beakers that the diluter can handle in one group then use the maximum number possible in that group and reallocate the remaining beakers by the procedure discussed below.

For purposes of planning the allocation of experimental effort among treatment groups, make the conservative assumption that $p_o=u_o$, $p_1=\ell_1$, ..., $p_I=\ell_I$. We calculate the sample sizes necessary to attain specified power when $p_i=\max(u_o+\Delta,\ell_i)$ $i=1,\ldots,I$. If $\ell_i<u_o+\Delta$ for all i, then an equal allocation plan is called for. The tighter the bounds around the p_i's are, the more unequal will be the allocation scheme.

Suppose type 1 error level α and power $1-\beta$ are desired. Let $\theta_o \equiv 2 \arcsin [u_o]^{1/2}$ and $\theta_i \equiv 2 \arcsin [\max(u_o+\Delta),\ell_i]^{1/2}$ $i=1,\ldots,I$. Let n_o, $n_1,\ldots, n_I$ denote the number of daphnids in each group and assume for discussion purposes that there is not any beaker-to-beaker variation. Let $Z_{1-\alpha}$, $Z_{1-\beta}$, denote the upper $1-\alpha$ and upper $1-\beta$ percentiles of the standard normal distribution respectively. Then the sample sizes must satisfy the relation

$$\frac{1}{n_i} + \frac{1}{n_o} = \frac{(\theta_1 - \theta_o)^2}{(Z_{1-\alpha} + Z_{1-\beta})^2} \quad i=1,\ldots,I$$

Consider a hypothetical example. Suppose I=5, $\Delta=.10$, $\alpha=.05$, $1-\beta=.90$. Then $Z_{.95}=1.645$, $Z_{.90}=1.282$, and

$0 \le p_o \le .05 \equiv u_o$	$u_o = .05$	$\theta_o=.45$
$\ell_1 \equiv 0 \le p_1 \le .10$	$\max(u_o+\Delta,\ell_1)=.15$	$\theta_1=.80$
$\ell_2 \equiv .10 \le p_2 \le .20$	$\max(u_o+\Delta,\ell_2)=.15$	$\theta_2=.80$
$\ell_3 \equiv .20 \le p_3 \le .30$	$\max(u_o+\Delta,\ell_3)=.20$	$\theta_3=.93$
$\ell_4 \equiv .40 \le p_4 \le .70$	$\max(u_o+\Delta,\ell_4)=.40$	$\theta_4=1.37$
$\ell_5 \equiv .80 \le p_5 \le 1.00$	$\max(u_o+\Delta,\ell_5)=.80$	$\theta_o=2.21.$

The group sample sizes n_i satisfy the relations

$$1/n_1 + 1/n_o = (.80-.45)^2/(1.645+1.282)^2 = .0143$$
$$1/n_2 + 1/n_o = .0143$$
$$1/n_3 + 1/n_o = .0269$$
$$1/n_4 + 1/n_o = .0988$$
$$1/n_5 + 1/n_o = .3616.$$

Thus n_o must be at least 70. The treatment allocations that satisfy the above relations and that approximately minimize the total numbers of daphnids required are:

$$n_o=175,\ n_1=117,\ n_2=117,\ n_3=48,\ n_4=11,\ n_5=2.$$

If a beaker contains 25 daphnids, this suggests having 7 control beakers, and 5,5,2,1,1 beakers allocated to the treatment groups (going from low to high concentrations).

XV. SPECIAL TOPICS. ANALYSIS OF TIME TRENDS IN PRODUCTIVITY. ANALYSIS OF TIME TO DEATH

A. INTRODUCTION

This section discusses several specialized topics. The analyses discussed here provide information concerning aspects of the data that were not considered in the previous sections.

Subsection B is concerned with the repeated measures analysis of time trends in productivity data. The analyses in Sections XI and XII pertain to the total numbers of offspring from the surviving daphnids and the variation in these total numbers among concentration groups. The analyses in this subsection consider the time patterns according to which these offspring were produced and the variation of these time patterns among concentration groups.

Subsection C is concerned with the analysis of time to death and the variation of the distributions of time to death among concentration groups. The analyses in Sections IX and X pertain to the numbers of surviving daphnids at various points in time. They do not utilize information about time to death. Such information increases the sensitivity of comparisons. The use of both parametric and nonparametric models is illustrated.

B. REPEATED MEASURES ANALYSIS OF TIME TRENDS IN PRODUCTIVITY

Goulden's data on isophorone are displayed in Figure I.8. The numbers of live offspring corresponding to days 2, 4, 6, 8, 11, 13, 15, 18, and 21 are shown for each group. Since there were no offspring and just one surviving daphnid in group 6, that group is deleted from comparisons of time trends in productivity. The data corresponding to the daphnid in group 2, beaker 1 are deleted from the comparisons because that daphnid died while on test. There were no offspring observed on days 2, 4, and 6 and just two daphnids produced offspring on day 8. Thus the comparisons of time trends in productivity were restricted to days 11 to 21, except for the two daphnids that produced offspring on day 8; the day 8 values were used for these daphnids.

The analyses of time trends in productivity require special analysis techniques since the numbers of offspring produced by the same daphnid at different points in time would be expected to be correlated. That is, some daphnids might produce relatively large numbers of offspring at each time point while others might produce relatively few offspring at each time point. Such data, consisting of responses over time for each

subject, are often referred to as repeated measures data or longitudinal data. A large number of techniques, of varying degrees of generality and complexity, are available for analyzing such data. The methods discussed in this section correspond to a relatively simple model and are readily applied.

Figure XV.1 displays the results of a repeated measures analysis of variance on the numbers of offspring produced on days 11, 13, 15, 18, and 21 by the surviving daphnids in groups 1 to 5 in the Goulden test on isophorone. The display was prepared with program P2V in the BMDP statistical computing system.

There are six panels in the analysis of variance table. The first panel corresponds to an analysis of variance on the average productivity of each daphnid across all five days. The three lines--MEAN, GROUP, and ERROR--correspond to a standard one way analysis of variance with five groups. This analysis of variance table is analagous to that discussed for the Goulden data in Section XI.B. The discussion there pertains to totals rather than averages and one observation is deleted form Group 1, but otherwise the results are analagous. Daphnid to daphnid random variability would be expected to be reflected in the average values across time and thus in the error sum of squares.

The second to fifth panels correspond to analyses of variance of various contrasts in the numbers of offspring across time. The notations N(1), N(2), N(3), N(4) correspond to the linear, quadratic, cubic, and quartic orthogonal polynomial components of the time trends in offspring. The top line in each panel coresponds to the overall effect across treatment groups, the second line corresponds to the interaction with groups, and the third line is the error term corresponding to that component. If the first mean square (e.g., N(1)) is statistically significant, this provides evidence of a significant overall polynomial trend component (e.g., linear component). If the second mean square (e.g., N(1)G) is statistically significant, this provides evidence that the values of the polynomial component vary from group to group. The manner in which they vary across groups needs to be studied further.

The sums of squares in the bottom panel are the sums of the corresponding sums of squares in the second to fifth panels. This panel represents a combined test of the presence of an overall time trend in numbers of offspring (NYNG effect) and the interaction of this time trend with group (NG effect). The pooling of the sums of squares for the orthogonal polynomial components assumes that these components are statistically independent and have the same variance. Two adjustments to the error degrees of freedom in the bottom panel account for departures from this assumption. The Greenhouse-Geisser and Huynh-Feldt adjustment factors are given at the bottom of the figure and their effects on the significance levels of the overall tests are shown to the right of the bottom panel.

The results in Figure XV.1 show very strong statistical evidence of group to group differences in average numbers of offspring on days 11 to 21. This is in direct correspondence with the results shown in Section XI.B. There are significant overall linear, quadratic, and cubic time trend components as well as a significant interaction between the linear component of time trend and concentration group.

Since the results in the second to sixth panels are based on contrasts across time, at least a portion of the daphnid to daphnid random variability would be expected to be eliminated from these components. Thus two error terms (denoted as 1 and 2 in the left most column of Figure XV.1) are shown. The first is used for comparing average values across time and the second is used for comparing contrasts within daphnids.

The analysis of variance calculations show the presence of time trends and the presence of group to group differences in the productivity responses; however they do not show the nature of these trends and differences. Additional analyses were carried out to characterize the nature of the differences and trends. For each daphnid, average responses over time and orthogonal polynomial contrasts were calculated. Let P_8, P_{11}, P_{13}, P_{15}, P_{18}, and P_{21} denote the numbers of offspring produced by a daphnid on days 8, 11, 13, 15, 18, and 21. respectively. For those daphnids with zero productivity on day 8 the orthogonal polynomial contrasts were defined as

$$AVG = (P_{11}+P_{13}+P_{15}+P_{18}+P_{21})/5^{1/2}$$

$$LIN = (-2P_{11}-P_{13}+P_{18}+2P_{21})/10^{1/2}$$

$$QUADR = (2P_{11}-P_{13}-2P_{15}-P_{18}+2P_{21})/14^{1/2}$$

$$CUBIC = (-P_{11}+2P_{13}-2P_{18}+P_{21})/10^{1/2}$$

The quartic contrast was not calculated because of its nonsignificance in the analysis of variance calculations in Figure XV.1. Contrasts analagous to those above (but including P_8) were calculated for the two daphnids that had offspring on day 8.

Histograms, summary statistics, and analysis of variance comparisons among groups are displayed for the AVG, LIN, QUADR, and CUBIC components in Figures XV.2 to XV.5, respectively. These displays were prepared with program P7D in the BMDP statistical computing system.

Figure XV.2 shows a quadratic like trend with concentration for the average numbers of offspring across time per daphnid. This trend is directly related to that displayed in Figure XII.14.

The overall mean values of the LIN, QUADR, and CUBIC components across all groups combined are substantially greater than their standard errors (see lower left portions of Figures XV.3, XV.4, XV.5). This reflects the significant linear, quadratic, and cubic main effects shown in Figure XV.1.

Figures XV.3 and XV.5 show significant group to group differences in the linear and cubic components, respectively. Figure XV.4 shows a significant difference among groups when the possibility of unequal variances is taken into account (Welch statistic) but not otherwise. The histograms in these three figures suggest linear trends with concentration or log concentration for each of the orthogonal polynomial components. The trend slopes are positive for the linear and quadratic components and negative for the cubic component.

Based on the results shown in Figures XV.1 to XV.5 polynomial regression functions were fitted to describe the trends in productivity across time and across concentration groups. As shown in Figure XV.1, separate error estimates are appropriate for analyzing the average (or total) productivity across time for each daphnid and for analyzing the contrasts that represent the time trends in productivity for each daphnid. These separate error estimates are referred to there as components 1 and 2, respectively.

Let P_i denote the number of offspring for a daphnid on the i-th day and let $\overline{P}$ denote the average number of offspring for that daphnid across all days. Then

$$P_i = \overline{P} + (P_i - \overline{P}) \quad .$$

We fit separate response curves to $\overline{P}$ and to $P_i - \overline{P}$.

The dose response curve fit to $\overline{P}$ is directly analagous to the dose response curves that were fitted in Section XII to the total productivity responses. In particular refer to Section XII.B and Figures XII.13, XII.14 for the fit to the Goulden data. Thus we need not repeat this discussion here.

The response curve fit to the $P_i - \overline{P}$ values reflects the trends in productivity both across time and across concentration groups. Although the $P_i - \overline{P}$ values within a daphnid are slightly negatively correlated, we treat these values as approximately independent for the purpose of the analysis illustrated below. This approximation can be refined by using generalized least squares or multivariate analysis techniques.

Let C denote concentration. Let D denote day and let $\overline{D}, s_d$ denote the mean and standard deviation of D. Define

$$\Delta P_i = P_i - \bar{P}$$

$$X = \log_{10}(1+C)$$

$$d = (D-\bar{D})/s_d$$

As suggested by the significance of the polynomial trend components in Figure XV.1 and by the natures of the concentration related trends displayed in Figures XV.3 to XV.5, we fit the following regression model to the trend data:

$$\Delta P_{ijk} = \beta_0+\beta_1 d_i+\beta_2 X_j+\beta_3 d_i X_j+\beta_4 d_i^2+\beta_5 d_i^2 X_j+\beta_6 d_i^3+\beta_7 d_i^3 X_j+\varepsilon_{ijk} \quad .$$

The indices i, j, k correspond to day, concentration, and replicate number, respectively. The term ε_{ijk} corresponds to the error term, which is assumed to be approximately independent with constant variance.

The results of this fit are shown in Table XV.6. All the estimated coefficients except those for the X_j and $d_i^2 X_j$ terms (X,IDY2X) are significant. The multiple R-square is just 0.21. Thus there is definite statistical evidence of the presence of time trends in the production of offspring and of the variation of these time trends with concentration; however most of the variation in numbers of offspring is random variation from daphnid to daphnid and is not explained by this systematic trend model.

We note from the signs of the regression coefficients that the concentration related trend in the linear time trend term (IDYX) is positive, that in the quadratic time trend term (IDY2X) is positive, and that in the cubic time trend term (IDY3X) is negative. This agrees with the concentration related trends observed in Figures XV.3 to XV.5.

Various inferences concerning the time trends in productivity and their relation to concentration can be based on the fitted regression model. For example it might be of interest to estimate the day associated with maximum productivity and its relation to concentration. The first time derivative of the regression function is zero at the time of maximum productivity and the second time derivative is negative. Let D_{max} denote the time of maximum productivity. Let $d_{max} = (D_{max}-\bar{D})/s_d$. Thus d_{max} satisfies the relations

$$3(\beta_6+\beta_7 X)d_{max}^2 + 2(\beta_4+\beta_5 X)d_{max} + (\beta_1+\beta_3 X) = 0$$

$$6(\beta_6+\beta_7 X)d_{max} + 2(\beta_4+\beta_5 X) < 0$$

Substitution of the estimated regression coefficients shown in Figure XV.6 into these relations, along with the values of X corresponding to the concentrations C, results in the following estimates of D_{max}.

Concentration, C	$\hat{D}_{max}$ (days)
0	12.7
10	12.8
50	13.2
100	14.0
150	15.7

The estimated times of maximum productivity are seen to increase with increasing concentration.

Approximate standard errors for these estimates can be calculated based on the variance-covariance matrix of the regression coefficients (not shown) and the delta method. The 1983 versions of the P3R and PAR nonlinear regression programs in the BMDP statistical computing system can directly calculate estimates of such nonlinear functions of the regression coefficients, along with associated standard error estimates and confidence intervals.

C. ANALYSIS OF TIME TO DEATH AND ITS RELATION TO CONCENTRATION

Sections IX and X contain analyses of the mortality responses. The statistical methods applied in those sections focus on one or on several study days and compare the mortality rates observed up to and including those days across groups. The methods and models applied in those sections do not utilize the specific time to death; just whether or not the death was prior to the day under consideration. These methods are based on the binomial or on the multinomial distribution.

Many investigators compile the (approximate) times to death. Analysis procedures exist that utilize such information. Such analysis procedures utilize more information than the binomial based procedures and thus might increase the sensitivity of inferences. We illustrate several such procedures in this subsection.

Figure XV.7 displays the observed times to death in Chapman's 21 day test on chromium. The columns in that figure contain the group number, the average measured concentration in each group, the day of death or of survival, the censor code, and the number of daphnids associated with that case (frequency). The censor code indicates whether the day indicated is associated with a death or with daphnids that survived to the end of the test. For daphnids that survived to the end of the test, the times to death are known just to the extent

that they exceed 21 days. Such data are referred to as right censored. The days associated with cases having censor code 1 represent numbers of days to death. The days associated with cases having censor code 2 represent survival beyond that number of days. Statistical procedures to analyze time to death data must account for such censoring.

Two aspects of the Chapman data shown in Figure XV.7 should be noted. First, the water control and the carrier control groups were combined for the purposes of the analyses in this subsection and are denoted as Group 1. There are 20 daphnids in this group; one died on day 10, one died on day 21, and 18 survived to the end of the test (i.e., beyond day 21). Secondly, the times to death shown are approximate. The beakers were examined on days 3, 5, 7, 10, 12, 14, 17, 19, and 21. The daphnids reported as dead at each time point actually died some time between the successive inspections. Thus a death indicated as day 14 could in fact have occurred on day 13 or 14; a death indicated as day 10 could in fact have occurred on days 8, 9, or 10. Such data are called interval censored; their values are known to lie in an interval. Statistical methods exist for analyzing interval censored data (Meeker and Duke,[33]). However for purposes of illustration we treat the times to death as if they occurred on the days indicated. If it is desired to utilize the times to death for statistical analysis purposes then the test beakers need to be inspected more than three times per week; they should probably be inspected daily.

Figure XV.8 contains comparisons of the 21 day mortality rates across treatment groups. Two tests are shown, the chi square test of homogeneity and the Cochran-Armitage test of linear trend. Both of these tests were discussed and illustrated in Section IX.

Both tests indicate strong statistical evidence of lack of homogeneity of 21 day mortality rates across treatment groups. The difference of the test statistics is 18.095-15.344=2.75], based on 6-1=5 d.f. The difference provides a test of nonlinear trends in mortality rates. Thus there is strong statistical evidence of a linear trend in mortality rates, but no statistical evidence of higher order trends.

The estimated standard errors of the 21 day mortality estimates are approximately as follows:

Group 1:	$N = 20,\ p = 0.1,\ s \doteq 0.067$
Groups 2,3:	$N = 10,\ p = 0.1,\ s \doteq 0.095$
Group 4:	$N = 10,\ p = 0.3,\ s \doteq 0.145$
Groups 5,6,7:	$N = 10,\ p = 0.5,\ s \doteq 0.158$

We now consider two approaches to the statistical analysis of the times to death; one is parametric and the other is nonparametric. We first consider the parametric analysis.

Preliminary probability plots (not shown) of the times to death (see Meeker and Duke,[33]) indicate that the distributions of times to death in each group can be approximated by the normal distribution. We fit a succession of regression models to describe the distributions of times to death and to compare these distributions among treatment groups.

The regression models are fitted to the data by maximum likelihood techniques using the CENSOR program (Meeker and Duke,[33]). CENSOR has the capability to accommodate censored values among the times to death, corresponding to the daphnids that survived to the end of the test. The models fitted to the mortality data assume that time to death is normally distributed with constant variance across groups. The mean time to failure depends on the treatment, in a manner specified for each model.

The first model fitted is an analyses of variance type model. It assumes different mean times to failure in each concentration group, with no assumptions concerning the form of the trend. This is an eight parameter model--a mean time to failure in each group and a common scale parameter. The results of fitting this model to the data are shown in Figure XV.9. The coefficient B_0 represents the estimated mean time to death (i.e., the 50th percentile) in Group 1, the control group. This is 35.59 days and is well beyond the end of the test. The coefficients B_1 to B_6 represent the differences between the estimated mean times to death in Groups 2 to 7 and that in Group 1. The estimated mean times to death in these groups are thus:

Group	Conc.	Mean Time to Death (Days)
2	13	35.59 - 3.16 = 32.43
3	29	35.59 - 3.16 = 32.43
4	66	35.59 - 10.30 = 25.29
5	132	35.59 - 18.00 = 17.59
6	294	35.59 - 14.16 = 21.43
7	655	35.59 - 14.54 = 21.05

The 95 percent confidence intervals on B_4, B_5, and B_6 do not contain 0. Thus the mean times to death are significantly lower in Groups 5, 6, and 7 than in the control group at the five percent level of significance, if adjustments are <u>not</u> made for simultaneously. After adjustment for simultaneous inferences by Bonferroni's techniques, only B_4 is significantly different from 0.

The second model fitted is a linear regression model. Let C denote the concentration. This model assumes that

$$\text{Mean Time to Death} = B_0 + B_1 \log_{10}(1+C)$$

Group	N	P_{reg}	Std Err_{reg}	P_{bin}	Std Err_{bin}
1	20	0.083	0.050	0.1	0.067
2	10	0.224	0.056	0.1	0.095
3	10	0.282	0.055	0.1	0.095
4	10	0.349	0.058	0.3	0.145
5	10	0.411	0.065	0.6	0.155
6	10	0.486	0.078	0.5	0.158
7	10	0.561	0.092	0.6	0.155

The two estimates of 21 day mortality are reasonably comparable for groups 1, 4, 6, and 7 and differ somewhat, due to smoothing, for groups 2, 3, and 4. The standard errors of the regression estimates are much reduced relative to those for the binomial estimates. For example for group 7 the estimated 21 day mortalities are similar whether the regression estimate or the standard binomial estimate is used. However the ratio of the standard errors of estimates is (0.155/0.092) = 1.685. This implies that it requires $(1.685)^2 = 2.84$ times as many daphnids to obtain comparable precision with this binomial based estimate as would be obtained with the regression estimate. The corresponding squared ratios for all but group 1 are close to or in excess of 3.

Thus utilizing the actual times to death and regression models to relate these times to death across *concentrations*, can result in substantially improved inference sensitivity as compared with the standard binomial theory estimates based on 21 day mortalities. Statistical procedures based on the times to death also provide mortality estimates at time points other than 21 days, such as 7 days or 14 days.

A normal probability plot of the residuals from this fit (not shown) was prepared and shows no evidence of departures from the model assumptions.

The previous techniques were based on a parametric regression model to describe the distribution of time to failure and to relate it to the level of test concentration. The previous models were based on the assumption that time to death was normally distributed, with constant variance across groups.

Cox[34] proposed a regression model that does not require specifying the form of the distribution of time to death. Let X denote a predictor varaiable; in our case X would be some function of concentration, such as log(1+C). Let $F_X(t)$ denote the cumulative distribution function of time to death of daphnids associated with predictor variable X and let $f_X(t)$ denote the corresponding probability density function. The hazard function, $h_X(t)$, is defined as

$$h_X(t) = f_X(t)/[1-F_X(t)] \quad .$$

with a common scale parameter in each group. The results of fitting this model to the data are shown in Figure XV.10. The coefficient B_0 represents the estimated intercept, and in this case coincides with the estimated mean time to death in the control group (37.37 days). The coefficient B_1 represents the regression slope; it is negative and statistically significant at the five percent level.

The maximum value of the log likelihood under the eight parameter model fit shown in Figure XV.9 is -115.5285. The maximum value of the log likelihood under the three parameter model fit shown in Figure XV.10 is -117.3918. Under the hypothesis that the linear regression is adequate to describe the trends in mean time to death, -2 times the difference of the log likelihoods is asymptotically distributed as chi square with 8-3=5 d.f. This provides an asymptotic test of the adequacy of the linear regression model. Namely

$$-2[-115.5285-(-117.3918)] = 3.7266 \quad .$$

Since this value is not significant according to chi square distribution with five degrees of freedom, we accept the hypothesis of the adequacy of a linear regression model.

An additional fit was carried out in which a single distribution of time to death was fitted to all seven groups. The maximum value of the log likelihood under this two parameter model is -123.6182. Under the hypothesis that $B_1 = 0$ in the previous linear regression model, -2 times the difference of the log likelihoods is asymptotically distributed as chi square with 3-2=1d.f. This provides an asymptotic test that $B_1 = 0$. Namely

$$-2[-117.3918-(-123.6182)] = 12.4528 \quad .$$

Since this value is highly significant according to the chi square distribution with one degree of freedom, we reject the hypothesis that $B_1 = 0$. We thus base estimates of mean time to death and of probability of death before various times, on the model fit shown in Figure XV.10.

Figure XV.11 contains estimates of the mean time to death for each concentration group and of the probabilities of dying by 7, 14, or 21 days. These estimates are based on the linear regression model fit. Each group can be identified by the transformed value of its concentration, $\log_{10}(1+C)$, shown under C_1.

The estimated probabilities of death by 21 days and their associated standard errors according to the linear regression model and according to the standard binomial estimates are:

Cox[34] proposed the proportional hazards regression model

$$h_x(t) = e^{\beta x} h_0(t)$$

where $h_0(t)$ is the hazard function associated with a reference distribution. The form of the $h_0(t)$ or its associated $F_0(t)$ do not need to be specified. In this sense the model is nonparametric. The proportional hazards model implies that the cumulative distribution functions of time to death are related by

$$1-F_x(t) = [1-F_0(t)]^{e^{\beta x}} \quad .$$

This model can be fitted to the data using a conditional maximum likelihood analysis suggested by Cox. This model was fitted to the times to death in Chapman's chromium data using the P2L program in the BMDP statistical computing system. The results of this fit are shown in Figure XV.12. The estimated value of β is $\hat{\beta} = 0.8427$ and $\hat{\beta}$ is statistically significant.

The fitted model is thus

$$1-F_x(t) = [1-F_0(t)]^{e^{0.8427X}}$$

where $X = \log_{10}(1+C)$. The survival probabilities shown in Figure XV.12 correspond to $X = \bar{X} = 1.4825$. Thus $1-F_{\bar{X}}(21) = 0.7654$ and

$$1-F_x(21) = (0.7654)^{2.3226^{(X-\bar{X})}} \quad .$$

Substituting the values of X corresponding to each test concentration yields

Group	Conc.	$X-\bar{X}$	$1-F_x(21)$	$F_x(21)$
1	0	-1.4825	0.926	0.074
2	13	-0.3364	0.818	0.182
3	29	-0.0054	0.766	0.234
4	66	0.3436	0.670	0.330
5	132	0.6414	0.632	0.368
6	294	0.9873	0.541	0.459
7	655	1.3344	0.439	0.561

These estimates are similar to those obtained from the parametric normal theory simple linear regression model.

ANALYSIS OF VARIANCE FOR 1-ST
DEPENDENT VARIABLE - P11 P13 P15 P18 P21

	SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.	GREENHOUSE GEISSER PROB.	HUYNH FELDT PROB.
	MEAN	30439.62043	1	30439.62043	481.67	0.0000		
	GROUP	5336.21317	4	1334.05329	21.11	.0000		
1	ERROR	1832.68095	29	63.19589				
	N(1)	616.86298	1	616.86298	13.49	.0010		
	N(1)G	691.42087	4	172.85522	3.78	.0136		
	ERROR	1326.47619	29	45.74056				
	N(2)	1003.08185	1	1003.08185	10.32	.0032		
	N(2)G	361.58994	4	90.39748	.93	.4605		
	ERROR	2820.03401	29	97.24255				
	N(3)	2271.26298	1	2271.26298	12.84	.0012		
	N(3)G	1119.41246	4	279.85312	1.58	.2054		
	ERROR	5128.19048	29	176.83415				
	N(4)	79.86838	1	79.86838	1.23	.2767		
	N(4)G	282.93359	4	70.73340	1.09	.3805		
	ERROR	1884.16599	29	64.97124				
	NYNG	3971.07619	4	992.76905	10.32	.0000	.0000	.0000
	NG	2455.35686	16	153.45980	1.60	.0807	.1203	.0982
2	ERROR	11158.86667	116	96.19713				

ERROR TERM — EPSILON FACTORS FOR DEGREES OF FREEDOM ADJUSTMENT

ERROR TERM	GREENHOUSE-GEISSER	HUYNH-FELDT
2	.6581	.8296

FIGURE XV.1 REPEATED MEASURES ANALYSIS OF VARIANCE OF NUMBERS OF OFFSPRING ON DAYS 11, 13, 15, 18, AND 21. GOULDEN ISOPHORONE-GROUP 6 DELETED. SURVIVING DAPHNIDS.

HISTOGRAM OF * AVG * (VARIABLE 10). CASES DIVIDED INTO GROUPS BASED ON VALUES OF * GROUP * (VARIABLE 1)

```
          0                   10                  50                  100                 150
          .....................+...................+...................+...................+.....................+
MIDPOINTS
  70.000)
  65.000)
  60.000)*
  55.000)
  50.000)                     *                   *
  45.000)                                         **
  40.000)                     M                   M*
  35.000)M***                 ****                *                   ***
  30.000)                                         *                   M**
  25.000)                                                             *
  20.000)**                                                                               *
  15.000)
  10.000)                                                                                 *
   5.000)                                                                                 M**
   0.000)                                                                                 **
```

GROUP MEANS ARE DENOTED BY M"S IF THEY COINCIDE WITH *"S, N"S OTHERWISE

MEAN	34.563	38.177	40.143	30.858	5.942
STD.DEV.	12.839	6.693	7.139	3.286	6.087
R.E.S.D.	10.453	7.054	7.389	3.632	6.079
S. E. M.	4.853	2.733	2.698	1.242	2.301
MAXIMUM	60.374	50.535	52.324	34.435	17.889
MINIMUM	21.913	33.094	29.963	25.938	0.000
SAMPLE SIZE	7	6	7	7	7

ALL GROUPS COMBINED (EXCEPT CASES WITH UNUSED VALUES FOR GROUP)

MEAN	29.694
STD.DEV.	14.696
R.E.S.D.	14.035
S. E. M.	2.520
MAXIMUM	60.374
MINIMUM	0.000
SAMPLE SIZE	34

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	5320.7052	4	1330.1763	21.36	.0000
WITHIN GROUPS	1805.9901	29	62.2755		
TOTAL	7126.6953	33			
LEVENE"S TEST FOR EQUAL VARIANCES		4, 29		.81	.5304
ONE-WAY ANALYSIS OF VARIANCE					
TEST STATISTICS FOR WITHIN-GROUP					
VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4, 14		28.84	.0000
BROWN-FORSYTHE		4, 17		21.60	.0000

FIGURE XV.2 HISTOGRAM, SUMMARY STATISTICS, AND ANALYSIS OF VARIANCE COMPARISONS OF AVG COMPONENT OF NUMBERS OF OFFSPRING ON DAYS 11 TO 21. GOULDEN ISOPHORONE - GROUP 6 DELETED. SURVIVING DAPHNIDS.

PAGE 5 GOULDEN PRODUCTIVITY DATA. DAYS 11 TO 21. GROUPS 1 TO 5. VARIATION OF POLYN COMPS AMONG GROUPS. PRDCTVTY.

HISTOGRAM OF * LIN * (VARIABLE 11). CASES DIVIDED INTO GROUPS BASED ON VALUES OF * GROUP * (VARIABLE 1)

```
          0                   10                  50                  100                 150
          +...................+...................+...................+...................+
MIDPOINTS
  18.000)
  15.000)
  12.000)                                         *
   9.000)                                         *                                    **
   6.000)                     *                                                        *
   3.000)                                                                              N
   0.000)                                                          **                  ****
  -3.000)***                  *                   M*              **
  -6.000)*                    M*                  **              M
  -9.000)M                                                        **
 -12.000)                     *
 -15.000)*                                       *
 -18.000)                     *
 -21.000)*
 -24.000)
```

GROUP MEANS ARE DENOTED BY M"S IF THEY COINCIDE WITH *"S, N"S OTHERWISE

MEAN	-8.583	-7.086	-1.627	-4.653	2.982
STD.DEV.	7.156	8.311	8.928	3.986	3.907
R.E.S.D.	7.304	7.986	9.208	4.386	4.735
S. E. M.	2.705	3.393	3.374	1.507	1.477
MAXIMUM	-2.530	5.692	10.752	.949	7.906
MINIMUM	-22.136	-18.341	-13.598	-9.803	-.949
SAMPLE SIZE	7	6	7	7	7

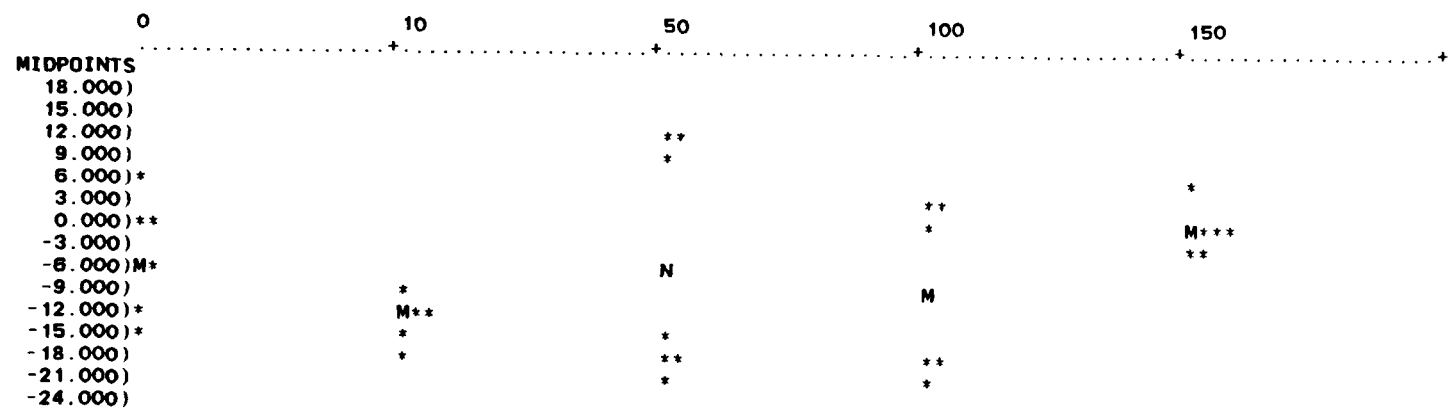
ALL GROUPS COMBINED (EXCEPT CASES WITH UNUSED VALUES FOR GROUP)	
MEAN	-3.697
STD.DEV.	7.593
R.E.S.D.	7.240
S. E. M.	1.302
MAXIMUM	10.752
MINIMUM	-22.136
SAMPLE SIZE	34

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	584.6760	4	146.1690	3.22	.0266
WITHIN GROUPS	1317.7883	29	45.4410		
TOTAL	1902.4643	33			
LEVENE"S TEST FOR EQUAL VARIANCES		4, 29		1.10	.3740
ONE-WAY ANALYSIS OF VARIANCE					
TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4, 14		4.94	.0107
BROWN-FORSYTHE		4, 21		3.15	.0355

FIGURE XV.3 HISTOGRAM, SUMMARY STATISTICS, AND ANALYSIS OF VARIANCE COMPARISONS OF LIN COMPONENT OF NUMBERS OF OFFSPRING ON DAYS 11 TO 21. GOULDEN ISOPHORONE - GROUP 6 DELETED. SURVIVING DAPHNIDS.

HISTOGRAM OF * QUADR * (VARIABLE 12). CASES DIVIDED INTO GROUPS BASED ON VALUES OF * GROUP * (VARIABLE 1)



GROUP MEANS ARE DENOTED BY M"S IF THEY COINCIDE WITH *"S, N"S OTHERWISE

MEAN	-4.734	-12.341	-5.179	-8.361	-.229
STD.DEV.	7.682	3.095	15.305	10.187	3.194
R.E.S.D.	7.974	3.144	18.864	11.976	2.658
S. E. M.	2.904	1.264	5.785	3.850	1.207
MAXIMUM	6.414	-7.751	12.657	3.207	6.414
MINIMUM	-15.768	-16.585	-19.510	-19.777	-3.742
SAMPLE SIZE	7	6	7	7	7

ALL GROUPS COMBINED (EXCEPT CASES WITH UNUSED VALUES FOR GROUP)	
MEAN	-5.987
STD.DEV.	9.568
R.E.S.D.	10.544
S. E. M.	1.641
MAXIMUM	12.657
MINIMUM	-19.777
SAMPLE SIZE	34

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	529.3337	4	132.3334	1.54	.2167
WITHIN GROUPS	2491.3913	29	85.9100		
TOTAL	3020.7250	33			
LEVENE"S TEST FOR EQUAL VARIANCES		4, 29		17.35	.0000
ONE-WAY ANALYSIS OF VARIANCE					
TEST STATISTICS FOR WITHIN-GROUP					
VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4, 14		10.63	.0004
BROWN-FORSYTHE		4, 15		1.60	.2262

FIGURE XV.4 HISTOGRAM, SUMMARY STATISTICS, AND ANALYSIS OF VARIANCE COMPARISONS OF **QUADR** COMPONENT OF NUMBERS OF OFFSPRING ON DAYS 11 TO 21. GOULDEN ISOPHORONE - GROUP 6 DELETED. SURVIVING DAPHNIDS.

HISTOGRAM OF * CUBIC * (VARIABLE 13). CASES DIVIDED INTO GROUPS BASED ON VALUES OF * GROUP * (VARIABLE 1)

```
          0                   10                  50                  100                 150
          .....................+...................+...................+...................+.....................+
MIDPOINTS
  32.000)
  28.000)
  24.000)****                 **                  **
  20.000)                     **
  16.000)N                    N                   *                   **
  12.000)**                   *
   8.000)                                         M                   M*
   4.000)*                                        *                   *                 **
   0.000)                                         *                   **                M****
  -4.000)                     *
  -8.000)
 -12.000)
 -16.000)
 -20.000)                                        *
 -24.000)
```

GROUP MEANS ARE DENOTED BY M"S IF THEY COINCIDE WITH *"S, N"S OTHERWISE

MEAN	17.393	15.981	7.404	7.047	1.039
STD.DEV.	8.083	10.582	15.372	7.047	2.488
R.E.S.D.	9.540	10.689	15.331	7.478	2.708
S. E. M.	3.055	4.320	5.810	2.663	.940
MAXIMUM	25.614	24.033	24.033	16.444	5.376
MINIMUM	4.743	-3.727	-20.871	-1.265	-1.581
SAMPLE SIZE	7	6	7	7	7

ALL GROUPS COMBINED (EXCEPT CASES WITH UNUSED VALUES FOR GROUP)	
MEAN	9.590
STD.DEV.	10.963
R.E.S.D.	11.685
S. E. M.	1.880
MAXIMUM	25.614
MINIMUM	-20.871
SAMPLE SIZE	34

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	1261.7436	4	315.4359	3.38	.0218
WITHIN GROUPS	2704.7957	29	93.2688		
TOTAL	3966.5393	33			
LEVENE"S TEST FOR EQUAL VARIANCES		4, 29		2.73	.0481
ONE-WAY ANALYSIS OF VARIANCE					
TEST STATISTICS FOR WITHIN-GROUP					
VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4, 13		8.22	.0016
BROWN-FORSYTHE		4, 17		3.35	.0338

FIGURE XV.5 HISTOGRAM, SUMMARY STATISTICS, AND ANALYSIS OF VARIANCE COMPARISONS OF **CUBIC** COMPONENT OF NUMBERS OF OFFSPRING ON DAYS 11 TO 21. GOULDEN ISOPHORONE - GROUP 6 DELETED. SURVIVING DAPHNIDS.

PAGE 4 TIME AND CONC TRENDS IN PRODUCTIVITY.

REGRESSION TITLE IS
TIME AND CONC TRENDS IN PRODUCTIVITY.

DEPENDENT VARIABLE. . . . . 5 DELPRD
TOLERANCE 0100
ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R .4606 STD. ERROR OF EST. 9.2220
MULTIPLE R-SQUARE .2122

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	3756.7756	7	536.6822	6.311	.0000
RESIDUAL	13947.3911	164	85.0451		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT		1.76553					
X	6	-.05279	1.37563	-.004	-.038	.9694	.40852
IDAY	10	-17.93176	3.48557	-1.906	-5.145	.0000	.03500
IDYX	11	8.29543	2.25345	1.413	3.681	.0003	.03263
IDY2	12	-3.33170	1.55139	-.344	-2.148	.0332	.18702
IDY2X	13	.97536	1.00066	.181	.975	.3311	.13883
IDY3	14	7.21343	1.65646	1.788	4.355	.0000	.02848
IDY3X	15	-3.42802	1.09971	-1.335	-3.117	.0022	.02620

FIGURE XV.6 REGRESSION MODEL TO DESCRIBE THE TIME TREND IN THE NUMBERS OF OFFSPRING ON DAYS 11 TO 21 AND ITS RELATION TO CONCENTRATION GROUP. GOULDEN ISOPHORONE - GROUP 6 DELETED. SURVIVING DAPHNIDS.

Group	Conc (µG/L)	Day	Censor Code	Freq
1	0	21	2	18
1	0	10	1	1
1	0	21	1	1
2	13	21	2	9
2	13	3	1	1
3	29	21	2	9
3	29	3	1	1
4	66	21	2	7
4	66	10	1	3
5	132	21	2	4
5	132	10	1	6
6	294	21	2	5
6	294	14	1	2
6	294	7	1	2
6	294	21	1	1
7	655	21	2	4
7	655	17	1	4
7	655	12	1	1
7	655	10	1	1

FIGURE XV.7 CHAPMAN-CHROMIUM. SURVIVAL DATA.

```
**************************
* TABLE PARAGRAPH   1  *
**************************

*****  OBSERVED FREQUENCY TABLE  1

VARIABLE     5   FREQ        USED AS COUNT VARIABLE.
                 ********

SURV                                          GROUP
------                                        ------
            CNTL       13       29       66       132      294      655      TOTAL
-------------------------------------------------------------------------------------
SURVIVE       18        9        9        7         4        5        4 I       56
DIE            2        1        1        3         6        5        6 I       24
-------------------------------------------------------------------------------I-------
TOTAL         20       10       10       10        10       10       10 I       80

        ALL CASES HAD COMPLETE DATA FOR THIS TABLE.


*****  PERCENTS OF COLUMN TOTALS -- TABLE  1

VARIABLE     5   FREQ        USED AS COUNT VARIABLE.
                 ********

SURV                                          GROUP
------                                        ------
            CNTL       13       29       66       132      294      655      TOTAL
-------------------------------------------------------------------------------------
SURVIVE     90.0     90.0     90.0     70.0      40.0     50.0     40.0 I     70.0
DIE         10.0     10.0     10.0     30.0      60.0     50.0     60.0 I     30.0
-------------------------------------------------------------------------------I-------
TOTAL      100.0    100.0    100.0    100.0     100.0    100.0    100.0 I    100.0

 MINIMUM ESTIMATED EXPECTED VALUE IS      3.00

 STATISTIC                      VALUE    D.F.   PROB.
 PEARSON CHISQUARE             18.095       6   .0060
 TEST FOR LINEAR TREND         15.344       1   .0001
```

FIGURE XV.8 CHAPMAN-CHROMIUM. CHI SQUARE TEST OF HOMOGENEITY ACROSS GROUPS AND COCHRAN-ARMITAGE TEST FOR LINEAR TREND. 21 DAY MORTALITY.

MAXIMUM LIKELIHOOD ESTIMATION COMPLETED

MAXIMUM VALUE OF THE LOGLIKELIHOOD IS -115.5285

PARAMETER ESTIMATES
FOR THE NORMAL DISTRIBUTION

TERM	ESTIMATE	STANDARD ERROR	95.00% CONFIDENCE LIMITS LOWER	UPPER
SCALE	11.4057	1.9294	8.1865	15.8909*
B 0	35.5891	5.0512	25.6868	45.4914
B 1	-3.1564	6.9394	-16.7605	10.4476
B 2	-3.1564	6.9394	-16.7605	10.4476
B 3	-10.2952	6.2077	-22.4648	1.8743
B 4	-18.0029	6.0844	-29.9307	-6.0751
B 5	-14.1600	6.1018	-26.1219	-2.1982
B 6	-14.5421	6.1099	-26.5199	-2.5643

* CONFIDENCE LIMITS FOR -SCALE- ASSUME THAT LOG(SCALE) IS NORMALLY DISTRIBUTED

VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED PARAMETERS

	SCALE	B 0	B 1	B 2	B 3	B 4
	B 5	B 6	B			
SCALE	.372E+01	.526E+01	-.171E+00	-.171E+00	-.228E+01	-.392E+01
	-.344E+01	-.387E+01				
B 0	.526E+01	.255E+02	-.183E+02	-.183E+02	-.213E+02	-.236E+02
	-.229E+02	-.235E+02				
B 1	-.171E+00	-.183E+02	.482E+02	.181E+02	.182E+02	.183E+02
	.182E+02	.183E+02				
B 2	-.171E+00	-.183E+02	.181E+02	.482E+02	.182E+02	.183E+02
	.182E+02	.183E+02				
B 3	-.228E+01	-.213E+02	.182E+02	.182E+02	.385E+02	.205E+02
	.202E+02	.205E+02				
B 4	-.392E+01	-.236E+02	.183E+02	.183E+02	.205E+02	.370E+02
	.217E+02	.222E+02				
B 5	-.344E+01	-.229E+02	.182E+02	.182E+02	.202E+02	.217E+02
	.372E+02	.217E+02				
B 6	-.387E+01	-.235E+02	.183E+02	.183E+02	.205E+02	.222E+02
	.217E+02	.373E+02				

FIGURE XV.9 CHAPMAN-CHROMIUM. RESULTS OF NORMAL DISTRIBUTION ANALYSIS OF VARIANCE MODEL FITTED TO TIMES TO DEATH. SEPARATE LOCATION PARAMETERS ASSUMED IN EACH GROUP. COMMON SCALE PARAMETER.

MAXIMUM LIKELIHOOD ESTIMATION COMPLETED

MAXIMUM VALUE OF THE LOGLIKELIHOOD IS -117.3918

PARAMETER ESTIMATES
FOR THE NORMAL DISTRIBUTION

TERM	ESTIMATE	STANDARD ERROR	95.00% CONFIDENCE LIMITS LOWER	UPPER
SCALE	11.8258	2.0012	8.4871	16.4779*
B 0	37.3724	4.9808	27.6080	47.1369
B 1	-6.4599	2.1135	-10.6032	-2.3167

* CONFIDENCE LIMITS FOR -SCALE- ASSUME THAT LOG(SCALE) IS NORMALLY DISTRIBUTED

VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED PARAMETERS

	SCALE	B 0	B 1	B
SCALE	.400E+01	.630E+01	-.175E+01	
B 0	.630E+01	.248E+02	-.943E+01	
B 1	-.175E+01	-.943E+01	.447E+01	

FIGURE XV.10 CHAPMAN-CHROMIUM. RESULTS OF NORMAL DISTRIBUTION SIMPLE LINEAR REGRESSION MODEL FITTED TO TIMES TO DEATH. COMMON SCALE PARAMETER ASSUMED IN EACH GROUP.

>FAIL PROBS R1 TIMES 7,14,21 Group 1

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
1	.0000	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 37.3724 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS		LOG ODDS L = LOG(F/(1-F))		95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER	LHAT	S(LHAT)	LOWER	UPPER
7.0000	.00511	.00541	.00064	.03977	-5.2715	1.0648	-7.3589	-3.1841
14.0000	.02406	.01886	.00508	.10642	-3.7031	.8035	-5.2783	-2.1278
21.0000	.08311	.05017	.02433	.24783	-2.4008	.6584	-3.6915	-1.1102

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

>FAIL PROBS R4 TIMES 7,14,21 Group 2

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
4	1.1461	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 29.9685 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS		LOG ODDS L = LOG(F/(1-F))		95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER	LHAT	S(LHAT)	LOWER	UPPER
7.0000	.02605	.01399	.00900	.07308	-3.6212	.5513	-4.7020	-2.5403
14.0000	.08846	.02958	.04515	.16610	-2.3326	.3668	-3.0517	-1.6135
21.0000	.22411	.05571	.13355	.35120	-1.2419	.3204	-1.8700	-.6138

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

FIGURE XV.11 CHAPMAN-CHROMIUM. ESTIMATES OF MEAN TIMES TO DEATH AND OF PROBABILITY OF DEATH BY 7, 14, AND 21 DAYS BASED ON THE LINEAR REGRESSION MODEL FIT SHOWN IN FIGURE XV.10.

>FAIL PROBS R6 TIMES 7,14,21

Group 3

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
6	1.4771	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 27.8303 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER
7.0000	.03908	.01755	.01601	.09230
14.0000	.12110	.03176	.07130	.19827
21.0000	.28178	.05521	.18687	.40110

LOG ODDS L = LOG(F/(1-F))

		95.00% CONFIDENCE LIMITS	
LHAT	S(LHAT)	LOWER	UPPER
-3.2022	.4674	-4.1185	-2.2859
-1.9820	.2984	-2.5670	-1.3971
-.9357	.2728	-1.4705	-.4009

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

>FAIL PROBS R8 TIMES 7,14,21

Group 4

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
8	1.8261	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 25.5761 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER
7.0000	.05811	.02280	.02655	.12250
14.0000	.16382	.03631	.10436	.24778
21.0000	.34939	.05822	.24531	.47013

LOG ODDS L = LOG(F/(1-F))

		95.00% CONFIDENCE LIMITS	
LHAT	S(LHAT)	LOWER	UPPER
-2.7855	.4165	-3.6020	-1.9690
-1.6301	.2650	-2.1497	-1.1105
-.6217	.2561	-1.1238	-.1196

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

FIGURE XV.11 (Continued)

>FAIL PROBS R10 TIMES 7,14,21

Group 5

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
10	2.1239	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 23.6525 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS		LOG ODDS L = LOG(F/(1-F))		95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER	LHAT	S(LHAT)	LOWER	UPPER
7.0000	.07954	.02952	.03773	.16000	-2.4486	.4031	-3.2389	-1.6582
14.0000	.20719	.04425	.13354	.30706	-1.3420	.2694	-1.8700	-.8139
21.0000	.41126	.06533	.29159	.54245	-.3587	.2698	-.8877	.1702

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

>FAIL PROBS R12 TIMES 7,14,21

Group 6

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS

	C 1	C
12	2.4698	

NORMAL DISTRIBUTION
LOCATION PARAMETER = 21.4175 AND SCALE PARAMETER = 11.8258

FAILURE PROBABILITY F

			95.00% CONFIDENCE LIMITS		LOG ODDS L = LOG(F/(1-F))		95.00% CONFIDENCE LIMITS	
Y	FHAT	S(FHAT)	LOWER	UPPER	LHAT	S(LHAT)	LOWER	UPPER
7.0000	.11139	.04122	.05250	.22093	-2.0766	.4164	-2.8929	-1.2603
14.0000	.26525	.05922	.16596	.39577	-1.0188	.3039	-1.6145	-.4231
21.0000	.48592	.07792	.33898	.63533	-.0563	.3119	-.6678	.5551

THESE LIMITS ASSUME
-LHAT- IS NORMALLY DISTRIBUTED

FIGURE XV.11 (Continued)

Group 7

```
>FAIL PROBS R16 TIMES 7,14,21

NORMAL DISTRIBUTION
FAILURE PROBABILITY ESTIMATES FOR TEST CONDITIONS
          C 1          C
   16     2.8169

NORMAL DISTRIBUTION
LOCATION PARAMETER =     19.1754 AND SCALE PARAMETER =     11.8258

                        FAILURE PROBABILITY F
                                    95.00% CONFIDENCE LIMITS
     Y          FHAT      S(FHAT)     LOWER       UPPER
   7.0000      .15161      .05804      .06871      .30208
  14.0000      .33083      .07930      .19676      .49944
  21.0000      .56131      .09218      .38052      .72717

                    THESE LIMITS ASSUME
                -LHAT- IS NORMALLY DISTRIBUTED

                        LOG ODDS L = LOG(F/(1-F))
                                    95.00% CONFIDENCE LIMITS
     LHAT       S(LHAT)     LOWER       UPPER
  -1.7220       .4513      -2.6067      -.8374
   -.7045       .3582      -1.4067      -.0022
    .2465       .3743       -.4874       .9803
```


FIGURE XV.11 (Continued)

INDEPENDENT VARIABLES
6 L10DS

LOG LIKELIHOOD = -96.0453
GLOBAL CHI-SQUARE = 10.95 D.F.= 1 P-VALUE = .0009

VARIABLE	COEFFICIENT	STANDARD ERROR	COEFF./S.E.	EXP(COEFF.)
6 L10DS	.8427	.2705	3.1151	2.3226

ESTIMATED ASYMPTOTIC COVARIANCE MATRIX

		L10DS 6
L10DS	6	.0732

ESTIMATED ASYMPTOTIC CORRELATION MATRIX

		L10DS 6
L10DS	6	1.000

FIGURE XV.12 CHAPMAN-CHROMIUM. RESULTS OF PROPORTIONAL HAZARDS REGRESSION MODEL FITTED TO TIMES TO DEATH.

TIME VARIABLE IS TIME

CASE LABEL	CASE NUMBER	SURVIVAL	STATUS	CUM DEATHS	CUM INCMPL	REMAIN AT RISK	KAPLAN MEIER SURVIVAL	PROPORTIONAL HAZARDS MODEL SURVIVAL EVALUATED FOR Z = ZBAR	HAZARD	CUM HAZARD	RESIDUAL
	40	3.00	DEAD	1	0	79		.9816		.0186	.0185
	30	3.00	DEAD	2	0	78	.9750	.9816	.0062	.0186	.0140
	69	7.00	DEAD	3	0	77		.9632		.0375	.0862
	68	7.00	DEAD	4	0	76	.9500	.9632	.0047	.0375	.0862
	80	10.00	DEAD	5	0	75		.8639		.1462	.4502
	60	10.00	DEAD	6	0	74		.8639		.1462	.2511
	59	10.00	DEAD	7	0	73		.8639		.1462	.2511
	58	10.00	DEAD	8	0	72		.8639		.1462	.2511
	57	10.00	DEAD	9	0	71		.8639		.1462	.2511
	56	10.00	DEAD	10	0	70		.8639		.1462	.2511
	55	10.00	DEAD	11	0	69		.8639		.1462	.2511
	50	10.00	DEAD	12	0	68		.8639		.1462	.1954
	49	10.00	DEAD	13	0	67		.8639		.1462	.1954
	48	10.00	DEAD	14	0	66		.8639		.1462	.1954
	19	10.00	DEAD	15	0	65	.8125	.8639	.0362	.1462	.0419
	79	12.00	DEAD	16	0	64	.8000	.8537	.0060	.1582	.4871
	67	14.00	DEAD	17	0	63		.8327		.1831	.4207
	66	14.00	DEAD	18	0	62	.7750	.8327	.0124	.1831	.4207
	78	17.00	DEAD	19	0	61		.7899		.2358	.7261
	77	17.00	DEAD	20	0	60		.7899		.2358	.7261
	76	17.00	DEAD	21	0	59		.7899		.2358	.7261
	75	17.00	DEAD	22	0	58	.7250	.7899	.0176	.2358	.7261
	70	21.00	DEAD	23	0	57		.7654		.2673	.6143
	20	21.00	DEAD	24	0	56	.7000	.7654	.0079	.2673	.0766
	74	21.00	CENSORED	24	1	55		.7654		.2673	.8230
	73	21.00	CENSORED	24	2	54		.7654		.2673	.8230
	72	21.00	CENSORED	24	3	53		.7654		.2673	.8230
	71	21.00	CENSORED	24	4	52		.7654		.2673	.8230
	65	21.00	CENSORED	24	5	51		.7654		.2673	.6143
	64	21.00	CENSORED	24	6	50		.7654		.2673	.6143
	63	21.00	CENSORED	24	7	49		.7654		.2673	.6143
	62	21.00	CENSORED	24	8	48		.7654		.2673	.6143
	61	21.00	CENSORED	24	9	47		.7654		.2673	.6143
	54	21.00	CENSORED	24	10	46		.7654		.2673	.4590
	53	21.00	CENSORED	24	11	45		.7654		.2673	.4590
	52	21.00	CENSORED	24	12	44		.7654		.2673	.4590
	51	21.00	CENSORED	24	13	43		.7654		.2673	.4590
	47	21.00	CENSORED	24	14	42		.7654		.2673	.3571
	46	21.00	CENSORED	24	15	41		.7654		.2673	.3571
	45	21.00	CENSORED	24	16	40		.7654		.2673	.3571
	44	21.00	CENSORED	24	17	39		.7654		.2673	.3571
	43	21.00	CENSORED	24	18	38		.7654		.2673	.3571
	42	21.00	CENSORED	24	19	37		.7654		.2673	.3571
	41	21.00	CENSORED	24	20	36		.7654		.2673	.3571
	39	21.00	CENSORED	24	21	35		.7654		.2673	.2661
	38	21.00	CENSORED	24	22	34		.7654		.2673	.2661
	37	21.00	CENSORED	24	23	33		.7654		.2673	.2661
	36	21.00	CENSORED	24	24	32		.7654		.2673	.2661
	35	21.00	CENSORED	24	25	31		.7654		.2673	.2661
	34	21.00	CENSORED	24	26	30		.7654		.2673	.2661
	33	21.00	CENSORED	24	27	29		.7654		.2673	.2661

FIGURE XV.12 (Continued)

CASE LABEL	CASE NUMBER	SURVIVAL	STATUS	CUM DEATHS	CUM INCMPL	REMAIN AT RISK	KAPLAN MEIER SURVIVAL	PROPORTIONAL HAZARDS MODEL SURVIVAL EVALUATED FOR Z = ZBAR	HAZARD	CUM HAZARD	RESIDUAL
	32	21.00	CENSORED	24	28	28		.7654		.2673	.2661
	31	21.00	CENSORED	24	29	27		.7654		.2673	.2661
	29	21.00	CENSORED	24	30	26		.7654		.2673	.2014
	28	21.00	CENSORED	24	31	25		.7654		.2673	.2014
	27	21.00	CENSORED	24	32	24		.7654		.2673	.2014
	26	21.00	CENSORED	24	33	23		.7654		.2673	.2014
	25	21.00	CENSORED	24	34	22		.7654		.2673	.2014
	24	21.00	CENSORED	24	35	21		.7654		.2673	.2014
	23	21.00	CENSORED	24	36	20		.7654		.2673	.2014
	22	21.00	CENSORED	24	37	19		.7654		.2673	.2014
	21	21.00	CENSORED	24	38	18		.7654		.2673	.2014
	18	21.00	CENSORED	24	39	17		.7654		.2673	.0766
	17	21.00	CENSORED	24	40	16		.7654		.2673	.0766
	16	21.00	CENSORED	24	41	15		.7654		.2673	.0766
	15	21.00	CENSORED	24	42	14		.7654		.2673	.0766
	14	21.00	CENSORED	24	43	13		.7654		.2673	.0766
	13	21.00	CENSORED	24	44	12		.7654		.2673	.0766
	12	21.00	CENSORED	24	45	11		.7654		.2673	.0766
	11	21.00	CENSORED	24	46	10		.7654		.2673	.0766
	10	21.00	CENSORED	24	47	9		.7654		.2673	.0766
	9	21.00	CENSORED	24	48	8		.7654		.2673	.0766
	8	21.00	CENSORED	24	49	7		.7654		.2673	.0766
	7	21.00	CENSORED	24	50	6		.7654		.2673	.0766
	6	21.00	CENSORED	24	51	5		.7654		.2673	.0766
	5	21.00	CENSORED	24	52	4		.7654		.2673	.0766
	4	21.00	CENSORED	24	53	3		.7654		.2673	.0766
	3	21.00	CENSORED	24	54	2		.7654		.2673	.0766
	2	21.00	CENSORED	24	55	1		.7654		.2673	.0766
	1	21.00	CENSORED	24	56	0		.7654		.2673	.0766

FIGURE XV.12 (Continued)

REFERENCES

[1] Feder, P. I. and Collins, W. J., Design and Analysis of Chronic Aquatic Tests of Toxicity", Draft Final Report, The Ohio State University Research Foundation, Contract No. DAMD17-79-C-9150, U.S. Army Medical Research and Development Command, May, 1981.

[2] Feder, P. I. and Willavize, Susan A., "EXAX2 -- A Computer Program to Compare Binomial Proportions, Program Description and Card Input Information", Technical Report No. 216, Department of Statistics, The Ohio State University, Columbus, Ohio, September, 1980.

[3] Williams, D. A., "The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity", Biometrics, Vol. 31, 1975, pp. 949-952.

[4] Finney, D. J., Probit Analysis, Third Edition, Cambridge University Press, London, 1971, pp. 175ff.

[5] Lehmann, E. L., Testing Statistical Hypotheses, 1959, John Wiley and Sons, Inc., Section 4.5.

[6] Lieberman, G. J. and Owen, D. B., Tables of the Hypergeometric Probability Distribution, 1961, Stanford University Press.

[7] Dixon, W. J. and Brown, M. B., Editors, BMDP Biomedical Computer Programs P-Series 1979, University of California Press, Berkeley, 1979.

[8] Dixon, W. J. and Massey, F. J., Jr., Introduction to Statistical Analysis, McGraw-Hill, New York, 1969, pp. 238, 240-243.

[9] Freund, J. E., Modern Elementary Statistics, Third Edition, Prentice-Hall, Inc., 1967, pp. 287-290.

[10] Helwig, J. T. and Council, K. A., Editors, SAS User's Guide, 1979 Edition, SAS Institute, Inc. Raleigh, 1979.

[11] Goodman, L. A. and Kruskal, W. H., "Measures of Association for Crossclassifications", Journal of the American Statistical Association, Vol. 49, December, 1954, pp. 732-764.

[12] Goodman, L. A. and Kruskal, W. H., "Measures of Association for Crossclassifications, IV: Simplification of Asymptotic Variances", Journal of the American Statistical Association, Vol. 67, No. 338, June, 1972, pp. 415-421.

[13] Proctor, C. H., "Relative Efficiencies of Some Measures of Association for Ordered Two-Way Contingency Tables Under Varying Intervalness of Measurement Errors", Proceedings of the Social Statistics Section, American Statistical Association, 1973, pp.372-379.

[14] Brown, M. B., and Benedetti, J. K., "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables", Journal of the American Statistical Association, Vol. 72, No. 358, June, 1977, pp. 309-315.

[15] Dunnett, C. W., "A Multiple Comparisons Procedure for Comparing Several Treatments With a Control", Journal of the American Statistical Association, Vol. 50, pp. 1096-1121.

[16] Dunnett, C. W., "New Tables for Multiple Comparisons With a Control", Biometrics, Vol. 20, 1964, pp. 482-491.

[17] Williams, D. A., "A Test for Differences Between Treatment Means When Several Dose Levels are Compared With a Zero Dose Control", Biometrics, Vol. 27, 1971, pp. 103-117.

[18] Williams, D. A., "The Comparison of Several Dose Levels With a Zero Dose Control", Biometrics, Vol. 28, 1972, pp. 519-531.

[19] Chew, V., Comparisons Among Means in an Analysis of Variance, Agricultural Research Service Publication ARS/H/6, October, 1977.

[20] Snedecor, G. W. and Cochran, W. G., Statistical Methods, Sixth Edition, Iowa State University Press, Ames, 1967, p. 246.

[21] Steel, R. G. D. and Torrie, J. H., Principles and Procedures of Statistics - A Biometrical Approach, Second Edition, McGraw-Hill, New York, 1980, p. 514.

[22] Fleiss, J. L., Statistical Methods for Rates and Proportion, Second Edition, John Wiley and Sons, 1981, pp. 145-146.

[23] Cochran, W. G., "Some Methods of Strengthening the Common χ^2 Tests", Biometrics, Vol. 10, 1954, pp. 417-451.

[24] Armitage, P., "Tests for Linear Trends in Proportions and Frequencies", Biometrics, Vol. 11, 1955, pp. 375-385.

[25] Jennrich, R. I. and Moore, R. H., "Maximum Likelihood Estimation by Means of Nonlinear Least Squares", in Proceedings of the Statistical Computing Section, American Statistical Association, 1975, pp. 57-65.

[26] Prentice, R. L., "A Generalization of the Probit and Logit Models for Dose Response Curves", Biometrics, Vol. 32, 1976, pp. 761-768.

[27] Cramér, H., Mathematical Methods of Statistics, Second Edition, Princeton University Press, 1965.

[28] Gelber, R. D., Lavin, P. T., Mehta, C. R., and Schoenfeld, D., "Statistical Methods in the Analysis of Aquatic Toxicity Studies", Unpublished notes, 1980.

[29] Krishnaiah, P. R., "Computations of Some Multivariate Distributions", Chapter 24 in Handbook of Statistics, Vol. 1, Analysis of Variance, Edited by P. R. Krishnaiah, North Holland Publishing Company, 1980, p. 745ff.

[30] Thomas, D. G. and Gart, J. J., "A Table of Exact Confidence Limits for Differences and Ratios of Two Proportions and Their Odds Ratios", Journal of the American Statistical Association, Vol. 72, 1977, pp. 73-76.

[31] Nelson, W., "Confidence Intervals for the Ratio of Two Poisson Means and Poisson Predictor Intervals", General Electric Company Research and Development Center TIS Report 69C118, February, 1969. IEEE Transactions on Reliability, 1970, Vol. R-19, pp. 42-49.

[32] Thomas, D. G., "Exact Confidence Limits for the Odds Ratio in a 2x2 Table", Applied Statistics, Vol. 20, 1971, pp. 105-110.

[33] Meeker, W. Q., Jr. and Duke, S. D., "CENSOR -- A User-Oriented Computer Program for Life Data Analysis User's Manual". Unpublished.

[34] Cox, D. R., "Regression Models and Life Tables", Jnl. Royal Statistic Society, Series B, Vol. 34, pp. 187-220.

APPENDIX AIV

NESTED ANALYSIS OF VARIANCE MODEL TO TEST FOR BEAKER TO BEAKER VARIATION WITHIN GROUPS

APPENDIX A.IV NESTED ANALYSIS OF VARIANCE MODEL TO TEST FOR BEAKER TO BEAKER VARIATION WITHIN GROUPS

In this appendix we specify the nested analysis of variance model that was used to test beaker to beaker variation within groups. This model is a nested model with beakers nested within groups and daphnids nested within beakers. Groups are fixed effects while beakers and daphnids are random effects.

Let $i \leftrightarrow$ group $\quad i = 1, \ldots, I.$
$j \leftrightarrow$ beaker $\quad j = 1, \ldots, J_i$
$k \leftrightarrow$ daphnids $\quad k = 1, \ldots, N_{ij}$

Let Y_{ijk} denote the length of the k-th daphnid within the j-th beaker of the i-th group. Then

$$Y_{ijk} = \mu + \alpha_i + b_{j(i)} + C_{k(ji)} + \varepsilon_{ijk}$$

$$\Sigma\alpha_i = 0, \; b_{j(i)} \overset{ind}{\sim} N(0, \sigma_b^2), \; C_{k(i)} \overset{ind}{\sim} N(0, \sigma_c^2), \; \varepsilon_{ijk} \overset{ind}{\sim} N(0, \sigma_\varepsilon^2) \quad .$$

The α_i's are the fixed group effects. The $b_{j(i)}$'s are the random beaker effects. The $C_{k(ji)}$'s are the random daphnids effects due to biological variation and to experimental variation. The ε_{ijk}'s are random measurement errors. Since we measure each daphnid just once, this source of variation cannot be separated from the biological and experimental variation and so we incorporate it into the daphnid to daphnid variation in subsequent discussion.

Let $\overline{Y}_{ij\cdot} \equiv \Sigma_k \, Y_{ijk}/N_{ij}$, $\quad N_{i+} \equiv \Sigma_j \cdot N_{ij}$, $\quad N_{++} \equiv \Sigma_i \, N_{i+}$,

$$\overline{\overline{Y}}_{i\cdot\cdot} \equiv \Sigma_j N_{ij} \, \overline{Y}_{ij\cdot}/N_{i+}, \; \overline{\overline{\overline{Y}}} \equiv \Sigma_i \, N_{i+} \, \overline{\overline{Y}}_{i\cdot\cdot}/N_{++} \quad .$$

Single, double, and triple bars over the b's and c's have analagous interpretations.

The analysis of variance table is given on the next page. In the special case of a balanced design (i.e. $J_i \equiv J$, $N_{ij} \equiv N$ for all i,j) the expected mean squares are

$$\sigma_c^2 + \frac{JN}{I-1} \Sigma_i \, \alpha_i^2 + N\sigma_b^2, \; \sigma_c^2 + N\sigma_b^2, \text{ and } \sigma_c^2 \quad .$$

Under the null hypothesis of no beaker to beaker variation within groups (i.e. $\sigma_b^2 = 0$) the F-ratio has a central F distribution with degrees of freedom $\Sigma_i(J_i-1)$ and $\Sigma_i \, \Sigma_j(N_{ij}-1)$. Under the alternative hypothesis it has a complicated, nonstandard distribution. However, in the balanced special case, the alternative distribution is $(1 + N\sigma_b^2/\sigma_c^2)$ times a central F distribution with degrees of freedom $I(J-1)$ and $IJ(N-1)$.

Source	d.f.	Sum of Squares	Expected Mean Square
Between Groups	$I-1$	$\Sigma_i N_{i+}(\bar{\bar{Y}}_{i..} - \bar{\bar{\bar{Y}}})^2$	$\sigma_c^2 + (I-1)^{-1} \Sigma_i N_{i+}(\alpha_i - \bar{\bar{\alpha}})^2 + (I-1)^{-1} E\Sigma_i N_{i+}(\bar{\bar{b}}_{.(i)} - \bar{\bar{\bar{b}}})^2$
Between Beakers Within Groups	$\Sigma_i(J_i-1)$	$\Sigma_i\Sigma_j N_{ij}(\bar{Y}_{ij} - \bar{\bar{Y}}_{i..})^2$	$\sigma_c^2 + E\Sigma_i\Sigma_j N_{ij}(b_{j(i)} - \bar{\bar{b}}_{.(i)})^2/\Sigma_i(J_i - 1)$
Within Beakers	$\Sigma_i\Sigma_j(N_{ij} - 1)$	$\Sigma_i\Sigma_j\Sigma_k(Y_{ijk} - \bar{Y}_{ij.})^2$	σ_c^2
Total	$N_{++}-1$	$\Sigma_i\Sigma_j\Sigma_k(Y_{ijk} - \bar{\bar{\bar{Y}}})^2$	

APPENDIX AV

ALTERNATIVE ADJUSTMENT PROCEDURE TO ACCOUNT FOR HETEROGENEITY OF MORTALITY RESPONSES IN THE PRESENCE OF UNEQUAL SAMPLE SIZES ACROSS BEAKERS WITHIN GROUPS OR ACROSS GROUPS

APPENDIX AV. ALTERNATIVE ADJUSTMENT PROCEDURE TO ACCOUNT FOR HETEROGENEITY OF MORTALITY RESPONSES IN THE PRESENCE OF UNEQUAL SAMPLE SIZES ACROSS BEAKERS WITHIN GROUPS OR ACROSS GROUPS

In subsection A we discussed a relatively simple procedure for adjusting sample sizes to reflect heterogeneity of mortality rates among beakers within groups. That procedure was based on an assumption of equal numbers of daphnids per beaker within each group. Such an equal sample size assumption is often reasonable in toxicity tests with daphnids because beakers are usually started with equal numbers of daphnids at the outset of the test. In particular this was the case in LeBlanc's Tests A and B and in Adams test with selenium. There are however some situations when the assumption of equal sample sizes might not be valid. For example in Goulden's test with isophorone there are ten beakers per group. Three of those beakers start the test with five daphnids each while the other seven start the test with just on daphnid each. If we wish to combine mortality results from the individually housed daphnids with those from the multiply housed daphnids then we may need to adjust for heterogeneity across beakers having unequal sample sizes. Similar situations arise when inferences for a given response are to be based on survivors up to a certain stage. For example in the fathead minnow early life stage tests the fry surviving for 30 days were examined for abnormality. Tank to tank variation in 30 day mortality within groups casues unequal sample sizes with respect to the fry abnormality response. A similar situation would arise in the analysis of data from toxicity tests with Daphnid if it is of interest to compare conditional mortality across groups. For example it might be of interest to compare mortality rates across groups based only the latter part of the life stage. Thus 21 day mortality might be compared across groups using responses only from survivors after 14 days. This would estimate the probability that a daphnid survives for 21 days given that it has survived for 14 days. Variation in 14 day survival among beakers within groups causes unequal sample sizes with respect to conditional 21 day survival.

Adjustments in the presence of such unequal sample sizes can be carried out with a full maximum likelihood fit based on the beta binomial model (Williams [3]). Namely within the i-th group, it is assumed that X_{ij} has a beta binomial distribution with parameters $(N_{ij}, \mu_i, \theta_i)$. The probability function of X_{ij} is then

$$P(X_{ij}=x) = \binom{N_{ij}}{x} \frac{Be\left(\frac{\mu_i}{\theta_i} + x,\ \frac{1-\mu_i}{\theta_i} + N_{ij} - x\right)}{Be\left(\frac{\mu_i}{\theta_i},\ \frac{1-\mu_i}{\theta_i}\right)} \qquad x = 0,1,..,\ N_{ij}$$

The parameter (μ_i, θ_i) can be estimated by maximum likelihood analysis. This would require specialized computer programs. The hypothesis H_0: $\theta_1 = \theta_2 = \ldots = \theta_I$ can be tested based on asymptotic maximum

likelihood theory. Adjustments can then be carried out using separate estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_I$ or a common estimate $\hat{\theta}$. The adjustment factor within the j-th beaker of the i-th group is then

$$\hat{K}_{ij} = \frac{1 + N_{ij}\,\hat{\theta}_i}{1 + \hat{\theta}_i} \quad \text{or} \quad K_{ij} = \frac{1 + N_{ij}\,\hat{\theta}}{1 + \hat{\theta}}$$

The responses within each beaker are then adjusted based on these factors.

If the $N_{ij} \equiv N_i$ are approximately constant within groups but vary across groups then $K_{ij} = K_i = (1+N_i\theta)/(1+\theta)$. Equal degrees of extrabinomial variations across groups would imply that $K_i = (1 + N_i\,\theta)/(1 + \theta)$. We obtain a pooled estimate of θ by calculating $\tilde{K}_i \equiv \hat{V}ar(P_{ij})/[\hat{\mu}_i\,(1-\hat{\mu}_i)/N_i]$ as before (bounding it between 1 and N_i) and calculating $\tilde{\theta}_i \equiv (\tilde{K}_i-1)/(N_i-\tilde{K}_i)$. We pool the $\tilde{\theta}_i$'s across groups by calculating the weighted average.

$$\hat{\bar{\theta}} = \frac{\Sigma_i\,(J_i-1)\,\tilde{\theta}_i}{\Sigma_i\,(J_i-1)}$$

The adjustment factors

$$K_i \equiv \frac{1 + N_i\,\hat{\bar{\theta}}}{1 + \hat{\bar{\theta}}}$$

are then used within each group.

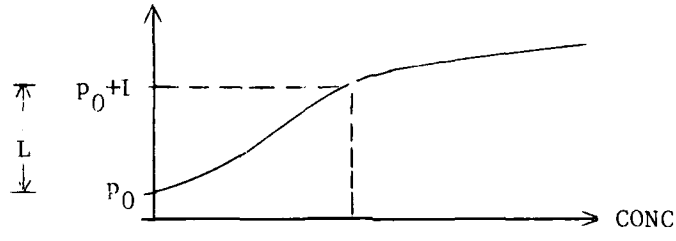
APPENDIX AX

CONFIDENCE INTERVAL ON CONCENTRATION THAT CORRESPONDS TO A GIVEN LEVEL OF INCREASE IN MORTALITY RATE OVER CONTROL GROUP RATE

APPENDIX AX. CONFIDENCE INTERVAL ON CONCENTRATION THAT CORRESPONDS TO A GIVEN LEVEL OF INCREASE IN MORTALITY RATE OVER CONTROL GROUP RATE

This appendix is patterned after Appendix AXV in Feder and Collins [1].

After we have fitted the probit model to the mortality data by non-linear regression we wish to calculate confidence bounds on the concentration, C_L, that results in an increase in mortality of L above the control group rate, p_0. Schematically,



Let $z \equiv \log_{10}(\text{conc}) - m, z_L \equiv \log_{10} C_L - m$. We fit a probit model in terms of z. We want a point estimate and confidence interval on C_L, such that $\Phi(\beta_0 + \beta_1 z_L) = L$. Let $\underset{\sim}{\theta} \equiv (p_0, \beta_0, \beta_1)$. We estimate $\underset{\sim}{\theta}$ by $\hat{\underset{\sim}{\theta}}$, the maximum likelihood estimate based on the probit model $p(\underset{\sim}{\theta}; \text{conc}) = p_0 + (1-p_0)\Phi(\beta_0 + \beta_1 z)$ using the program BMDPAR. Thus z_L satisfies the equation

$$\hat{\beta}_0 + \hat{\beta}_1 z_L = \Phi^{-1}(L) \equiv f_L$$

or

$$\hat{z}_L = (f_L - \hat{\beta}_0)/\hat{\beta}_1 \equiv g(\hat{\beta}_0, \hat{\beta}_1; L)$$

We construct a confidence interval on z_L based on the estimate $\hat{z}_L$ and the delta method (Cramér [27], pp. 366-367). Asymptotically, as sample size increases, the function $g(\hat{\beta}_0, \hat{\beta}_1; L)$ can be approximated by a first order Taylor expansion about β_0 and β_1.

$$\hat{z}_L \tilde{=} z_L + (\hat{\beta}_0 - \beta_0)\, \partial g/\partial \beta_0 + (\hat{\beta}_1 - \beta_1)\, \partial g/\partial \beta_1$$

Under certain standard regularity assumptions, the asymptotic distribution of (appropriate standardizations of) $\hat{z}_L$ can be shown to be the same as that of its Taylor approximation, namely normal with mean z_L and variance

$$\hat{\text{Var}}(\hat{z}_L) \cong (\hat{\partial g}/\partial\beta_0,\ \hat{\partial g}/\partial\beta_1)\ \hat{\Sigma} \begin{pmatrix} \hat{\partial g}/\partial\beta_0 \\ \hat{\partial g}/\partial\beta_1 \end{pmatrix}$$

where $\hat{\Sigma}$ is the estimated variance-covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\partial g}/\partial\beta_0$, $\hat{\partial g}/\partial\beta_1$, are the first derivatives of g evaluated at $\hat{\beta}_0$ and $\hat{\beta}_1$. The matrix $\hat{\Sigma}$ is determined from the BMDPAR output. The derivatives of g are

$$\partial g/\partial\beta_0 = -1/\beta_1$$

$$\partial g/\partial\beta_1 = (\beta_0 - f_L)/\beta_1^2$$

The asymptotic $1-\alpha$ confidence interval on z_L is thus

$$z_L \in \hat{z}_L \pm \xi_{\alpha/2}\ [\hat{\text{Var}}\ (\hat{z}_L)]^{1/2}$$

where $\xi_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution.

These calculations can easily be programmed on a computer or calculator.

APPENDIX AXII.1

THEORY UNDERLYING POINT AND CONFIDENCE INTERVAL ESTIMATION OF CONCENTRATIONS ASSOCIATED WITH SPECIFIED REDUCTIONS IN AVERAGE REPRODUCTION OR LENGTH RELATIVE TO THE CONTROL GROUP

APPENDIX AXII.1 THEORY UNDERLYING POINT AND CONFIDENCE INTERVAL ESTIMATION OF CONCENTRATIONS ASSOCIATED WITH SPECIFIED REDUCTIONS IN AVERAGE REPRODUCTION OR LENGTH RELATIVE TO THE CONTROL GROUP

Let $x \equiv \log_{10}$ (concentration), $v = (x-m)/s$, I=indicator of treatment groups. We fit the regression model

$$Y=\mu+\beta_o I+\beta_1 Ix+\varepsilon$$

or

$$Y=\mu+\beta_o I+\beta_1 Iv+\beta_2 Iv^2+\varepsilon$$

where ε represents the random variation, assumed to be independently distributed with mean 0 and variance σ^2. Let c denote a specified incremental response from the control group average. We wish to estimate the value of v such that

$$\beta_o+\beta_1 v=c$$

or

$$\beta_o+\beta_1 v+\beta_2 v^2=c$$

and place confidence intervals on this value. The starting point for these inferences is the output from the regression analysis program which provides estimates of model parameters $(\hat{\beta}_o,\hat{\beta}_1)$ or $(\hat{\beta}_o,\hat{\beta}_1,\hat{\beta}_2)$ and the estimated variance-covariance matrix of these parameters, which we denote as $\hat{\underline{\underline{\Sigma}}}$. Let v denote the residual degrees of freedom. We consider, in turn, the straight line and the quadratic cases.

Straight Line Case

We solve the equation

$$(\beta_o-c)+\beta_1 v=0$$

thus

$$v_c = -(\beta_o-c)/\beta_1$$

The point estimate, $\hat{v}_c$, of v_c is thus

$$\hat{v}_c = -(\hat{\beta}_o - c)/\hat{\beta}_1$$

We construct an approximate confidence interval on v_c by use of the delta method (sometimes referred to as propagation of errors). Let $v_c \equiv g(\beta_0,\beta_1)$. Then $\hat{v}_c = g(\hat{\beta}_0,\hat{\beta}_1)$. Expand $g(\beta_0,\beta_1)$ in a first order Taylor expansion about (β_o,β_1). Thus

$$\hat{v}_c \equiv g(\hat{\beta}_o,\hat{\beta}_1) = g(\beta_o,\beta_1) + (\hat{\beta}_o - \beta_o)\partial g/\partial\beta_o + (\hat{\beta}_1 - \beta_1)\partial g/\partial\beta_1 + \text{remainder}$$

The remainder becomes small asymptotically as the sample size gets large. Approximating $\hat{v}_c$ by the first order terms in the expansion, we obtain the result (loosely stated) that $\hat{v}_c$ is asymptotically normally distributed with mean v_c and variance

$$\text{Var}(\hat{v}_c) \doteq (\partial g/\partial\beta_o, \partial g/\partial\beta_1)\ddagger(\partial g/\partial\beta_o, \partial g/\partial\beta_1)'$$

We approximate this asymptotic variance by substituting $\hat{\ddagger}$ for $\ddagger$ and by substituting $(\hat{\beta}_o,\hat{\beta}_1)$ for (β_o,β_1) in the expressions for $\partial g/\partial\beta_o, \partial g/\partial\beta_1$. Let $\hat{\text{Var}}(\hat{v}_c)$ denote this estimated variance.

To complete the characterization of the asymptotic distribution of $\hat{v}_c$, we need to specify the functional forms of the derivatives. These are

$$\partial g/\partial\beta_o = -1/\beta_1 \qquad \partial g/\partial\beta_1 = (\beta_o - c)/\beta_1^2$$

An approximate $1-\alpha$ confidence interval on v_c is

$$v_c \varepsilon \hat{v}_c \pm t(1-\alpha/2;\nu)[\hat{\text{Var}}(\hat{v}_c)]^{1/2} \equiv (\hat{v}_\ell, \hat{v}_u)$$

where $t(1-\alpha/2;\nu)$ is the upper $\alpha/2$ point of the t distribution with ν degrees of freedom.

To translate this point estimate and confidence interval to an estimate and confidence interval directly on concentration, we use the relation

$$\text{conc} = 10^{m+sv}$$

thus

$$\hat{\text{conc}}_c = 10^{m+s\hat{v}_c}, \quad \hat{\text{conc}}_\ell = 10^{m+s\hat{v}_\ell}, \quad \hat{\text{conc}}_u = 10^{m+s\hat{v}_u}$$

These are the inferential values of interest.

Quadratic Case

This is conceptually very similar to the straight line case however there are several technical complications due to the greater complexity of the model.

We solve the equation

$$(\beta_0-c)+\beta_1 v+\beta_2 v^2=0$$

There are two roots to the equation:

$$v_c = \frac{-\beta_1 \pm [\beta_1^2-4\beta_2(\beta_0-c)]^{1/2}}{2\beta_2} \qquad \text{if } \beta_2 \neq 0$$

$$v_c = \frac{-(\beta_0-c)}{\beta_1} \qquad \text{if } \beta_2 = 0$$

We considered the $\beta_2=0$ case in the discussion of the straight line case. We therefore assume below that $\beta_2 \neq 0$.

The point estimate, $\hat{v}_c$, of v_c is

$$\hat{v}_c = \frac{-\hat{\beta}_1 \pm [\hat{\beta}_1^2-4\hat{\beta}_2(\hat{\beta}_0-c)]^{1/2}}{2\hat{\beta}_2} \qquad \text{if } \hat{\beta}_2 \neq 0,\ \hat{\beta}_1^2-4\hat{\beta}_2(\hat{\beta}_0-c) \geq 0$$

We will generally be interested in the larger root. Which root is largest depends on the values of the coefficients. Rather than attempt to choose the appropriate root a priori, we will calculate point and confidence interval estimates of both roots and then will choose which is most appropriate from the context of the problem.

We construct an approximate confidence interval on v_c by use of the delta method, much as we did in the straight line case. Let

$$v_- \equiv f_-(\beta_0,\beta_1,\beta_2) \equiv \frac{-\beta_1-[\beta_1^2-4\beta_2(\beta_0-c)]^{1/2}}{2\beta_2}$$

$$v_+ \equiv f_+(\beta_0,\beta_1,\beta_2) \equiv \frac{-\beta_1+[\beta_1^2-4\beta_2(\beta_0-c)]^{1/2}}{2\beta_2}$$

Define $V \equiv [\beta_1^2 - 4(\beta_0 - c)\beta_2]^{1/2}$. Expand $f_-(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$, $f_+(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ in first order Taylor expansions about $(\beta_0, \beta_1, \beta_2)$. Thus

$$\hat{v}_+ \equiv f_+(\beta_0, \beta_1, \beta_2) = f_+(\beta_0, \beta_1, \beta_2) + (\hat{\beta}_0 - \beta_0)\partial f_+/\partial\beta_0 + (\hat{\beta}_1 - \beta_1)\partial f_+/\partial\beta_1 + (\hat{\beta}_2 - \beta_2)\partial f_+/\partial\beta_2 + \text{remainder}$$

$$\hat{v}_- \equiv f_-(\beta_0, \beta_1, \beta_2) = f_-(\beta_0, \beta_1, \beta_2) + (\hat{\beta}_0 - \beta_0)\partial f_-/\partial\beta_0 + (\hat{\beta}_1 - \beta_1)\partial f_-/\partial\beta_1 + (\hat{\beta}_2 - \beta_2)\partial f_-/\partial\beta_2 + \text{remainder}$$

Approximating $\hat{v}_+$, $\hat{v}_-$ by the first order terms in the expansions, we obtain the result that $\hat{v}_+$, $\hat{v}_-$ are asymptotically normally distributed with means $\hat{v}_+$, $\hat{v}_-$ and variances

$$\mathrm{Var}(\hat{v}_-) \doteq (\partial f_-/\partial\beta_0,\ \partial f_-/\partial\beta_1,\ \partial f_-/\partial\beta_2)\ \ddagger \begin{pmatrix} \partial f_-/\partial\beta_0 \\ \partial f_-/\partial\beta_1 \\ \partial f_-/\partial\beta_2 \end{pmatrix}$$

$$\mathrm{Var}(\hat{v}_+) = (\partial f_+/\partial\beta_0,\ \partial f_+/\partial\beta_1,\ \partial f_+/\partial\beta_2)\ \ddagger \begin{pmatrix} \partial f_+/\partial\beta_0 \\ \partial f_+/\partial\beta_1 \\ \partial f_+/\partial\beta_2 \end{pmatrix}$$

We approximate these asymptotic variances by substituting $\hat{\ddagger}$ for $\ddagger$ and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ for $(\beta_0, \beta_1, \beta_2)$ in the expressions for the derivatives. Let $\widehat{\mathrm{Var}}(\hat{v}_-)$, $\widehat{\mathrm{Var}}(\hat{v}_+)$ denote these estimated variances. Approximate $1-\alpha$ confidence intervals on $\hat{v}_-, \hat{v}_+$ are

$$v_- \varepsilon \hat{v}_- \pm t(1-\alpha/2;\nu)[\widehat{\mathrm{Var}}(\hat{v}_-)]^{1/2} \equiv (\hat{v}_{\ell-}, \hat{v}_{u-})$$

$$v_+ \varepsilon \hat{v}_+ \pm t(1-\alpha/2;\nu)[\widehat{\mathrm{Var}}(\hat{v}_+)]^{1/2} \equiv (\hat{v}_{\ell+}, \hat{v}_{u+})$$

where $t(1-\alpha/2;\nu)$ is the upper $\alpha/2$ point of the t distribution with ν degrees of freedom. We then transform these estimates and confidence intervals in the usual manner to obtain estimates and confidence intervals directly on concentration. Namely we use the relation

$$conc=10^{m+sv}$$

To complete the characterization of the asymptotic distributions of $\hat{v}_-$ $\hat{v}_+$, we need to specify the functional forms of the derivatives.

These are

$$\partial f_-/\partial\beta_0 = V^{-1} \qquad \partial f_-/\partial\beta_1 = \frac{-1-V^{-1}\beta_1}{2\beta_2} \qquad \partial f_-/\partial\beta_2 = \frac{V^{-1}(\beta_0-c)}{\beta_2} + \frac{\beta_1+V}{2\beta_2^2}$$

$$\partial f_+/\partial\beta_0 = -V^{-1} \qquad \partial f_+/\partial\beta_1 = \frac{-1+V^{-1}\beta_1}{2\beta_2} \qquad \partial f_+/\partial\beta_2 = \frac{-V^{-1}(\beta_0-c)}{\beta_2} + \frac{\beta_1-V}{2\beta_2^2}$$

APPENDIX AXII.2

CONFINT--A COMPUTER PROGRAM TO CALCULATE POINT AND CONFIDENCE INTERVAL ESTIMATES OF CONCENTRATIONS ASSOCIATED WITH SPECIFIED REDUCTIONS IN AVERAGE REPRODUCTION OR LENGTH RELATIVE TO THE CONTROL GROUP

APPENDIX AXII.2 CONFINT--A COMPUTER PROGRAM TO CALCULATE POINT AND CONFIDENCE INTERVAL ESTIMATES OF CONCENTRATIONS ASSOCIATED WITH SPECIFIED REDUCTIONS IN AVERAGE REPRODUCTION OR LENGTH RELATIVE TO THE CONTROL GROUP

In Appendix AXII.1 we discussed the theory underlying the construction of point and confidence interval estimates of concentrations associated with specified reductions in average length or reproduction by means of the delta method. In this appendix we describe the features and use of a computer program to implement that theory. The program was written by Claire Matthews. Applications of the program are illustrated in the body of the section.

General Description

CONFINT is a FORTRAN computer program which calculates a point estimate and 95 percent confidence interval for the concentration producing a specified change, c, in the average response relative to the control group average. The program assumes that the response parameter of interest (y) has a linear or quadratic dose-response relationship with logarithmic concentration (x), among the treatment groups. Therefore the underlying model is $y=\mu$ for the control group data, and

$$y=f(x)=\mu+\beta_0+\beta_1 x \qquad (1)$$

or

$$y=f(x)=\mu+\beta_0+\beta_1 x+\beta_2 x^2 \qquad (2)$$

for treatment group data, depending on whether a linear or quadratic model is appropriate. The program calculates the root(s) of the equation $f(x)=\mu+c$, where c represents the extent of change relative to the control group response, as specified by the user.

Estimates of the coefficients of the above model are input requirements to CONFINT. The most straightforward way to obtain these estimates is to first run a standard multiple linear regression analysis, using any standard statistical package. To incorporate data from both the control group and the treatment groups in the same regression run, define an indicator variable I where I=1 for treatment groups and 0 for control groups. Then create independent variables I, Ix, and Ix^2, and use the regression program to fit the model

$$y=\mu+\beta_0 I+\beta_1 Ix+\varepsilon \qquad (1A)$$

or

$$y=\mu+\beta_0 I+\beta_1 Ix+\beta_2 Ix^2+\varepsilon \qquad (2A)$$

This yield values of the estimates.

In addition to the regression coefficient estimates, CONFINT requires the estimated variance covariance matrix of the $\{\hat{\beta}_i\}$ coefficients. The

SPSS regression program e.g., can print out this matrix through one of its regression output options.

It is often desirable, from a numerical analysis standpoint, to rescale the x values to v=(x-m)/s in Equations (1A) and (2A) in order to reduce correlations among the regression coefficients and to scale the regression coefficients and variance covariance matrix elements so that they are all about the same order of magnitude and that none are so large or so small that significant digits are lost in the regression output. Any values can be chosen for m and s. However they must be recorded, since they must also be specified as input parameters to CONFINT. A convenient choice of m and s, although not the only one, is the mean and standard deviation, respectively, of the x-values of the treatment group observations.

Program Input Specifications

The input specifications for each problem, read into CONFINT on (input unit) TAPE5, consist of a set of six or seven cards depending on whether a linear or quadratic model is used. For each set of input parameters the program will calculate one set of point estimates and 95 percent confidence intervals corresponding to the user-specified change, c, from the mean control response.

The input cards are to be arranged as follows:

Card	Columns	Format*	Variable
1	1-80	10A8	any title occupying any of the 80 columns
2	1- 5	I5	no. of degrees of freedom for error in the regression fit
	6-10	(blank)	
	11-20	F10.0	regression estimate for μ
	21-30	F10.0	regression estimate for β_o
	31-40	F10.0	regression estimate for β_1
	41-50	F10.0	regression estimate for β_2--leave blank if a straight line model is used
3	1- 5	I5	ISCALE=0 if no scale charge was used for x = 1 if a scale change of the form v=(x-m)/s was used
	6-10	(blank)	
	11-20	F10.0	value used for m } leave blank if
	21-30	F10.0	value used for s } ISCALE = 0.
4	1- 5	I5	IUNIT = 1 if c is given in absolute units; = 0 if c is given as a decimal (proportion relative to the control mean).
	6-10	(blank)	
	11-20	F10.0	value for c, which can be either positive or negative. For example, if one is

			interested in the concentration associated with a 50 percent reduction from control group average (i.e., an EC50), use IUNIT=0 and C=-0.50
5	1-10	F10.0	variance $(\hat{\beta}_0)$
6	1-10	F10.0	covariance $(\hat{\beta}_0, \hat{\beta}_1)$
	11-20	F10.0	variance $(\hat{\beta}_1)$
7**	1-10	F10.0	covariance $(\hat{\beta}_0, \hat{\beta}_2)$
	11-20	F10.0	covariance $(\hat{\beta}_1, \hat{\beta}_2)$
	21-30	F10.0	variance $(\hat{\beta}_2)$
1	1-80	10A8	Title card for a new problem if necessary.
etc.	etc.	etc.	Cards 1-7 can be repeated as many times as desired for any number of problems. An end-of-file card should follow the last problem.
7			
(or end-of-file card)			

---

*Note: An "F10.0" format indicates that the user should enter a floating point number in the field of 10 columns (not necessarily right-justified); a decimal point must be punched, although it can appear anywhere in the number. Variables having the "I5" format, however must be entered as right-justified integers ending in column 5.

**Card 7 should be omitted if a linear model is used. Note that cards 5-7 are set up to contain the lower diagonal portion of the variance-covariance matrix for the $\{\hat{\beta}_i\}$ regression coefficients.

Program Output

A separate output is produced for each problem inputted to CONFINT. Each output first includes a printout of the input specifications, including the model chosen by the user, regression coefficients, rescaling transformation (if any), error degrees of freedom, and the estimated covariance matrix of the $\{\hat{\beta}_i\}$ regression coefficients. The value of c is printed in absolute y units; if c was read in as a decimal using relative units, then the program recalculates c as $c \cdot \hat{\mu}$ before printing it back.

The program reports calculations for both roots of the equation $f(x)=\mu+c$ if a quadratic model is used and if two roots exist. (If no roots exist, the program prints out a message to this effect before going on to the next problem.) The partial derivatives of the polynomial roots with respect to the β_i's are printed out. From these derivatives and the estimated variance covariance matrix, the estimated variances of the roots are calculated and printed out.

Finally, the program prints out the point estimates of the roots and their associated 95 percent confidence intervals, in three different scaling systems:

(1) in rescaled units, if x was rescaled to $v \equiv (x-m)/s$.

(2) in log-concentration units (regular x units), and

(3) in raw concentrations, where antilogarithms are taken of all the values reported in (2).

In the event that x already represented raw concentrations, the values reported for (3) should be ignored.

It should be noted that whenever the quadratic model yields two roots, the smaller of the two roots may be much lower than the lowest treatment concentration. It is calculated and printed out for mathematical completeness, but it should probably be disregarded ecause it may be totally unrealistic in the physical context of the experiment. The specific root to be used depends on the context of the problem, but it will usually be the larger root.

APPENDIX AXIII

THEORY UNDERLYING THE CONSTRUCTION OF CONFIDENCE INTERVALS ON TREATMENT GROUP-CONTROL GROUP MORTALITY RATE DIFFERENCES BASED ON THE RESULTS OF THREE PARAMETER PROBIT MODEL REGRESSION FITS

APPENDIX AXIII. THEORY UNDERLYING THE CONSTRUCTION OF CONFIDENCE INTERVALS ON TREATMENT GROUP-CONTROL GROUP MORTALITY RATE DIFFERENCES BASED ON THE RESULTS OF THREE PARAMETER PROBIT MODEL REGRESSION FITS

Let

$$p(\text{conc}) = p_o + (1-p_o)\Phi(\beta_o + \beta_1(z-m))$$

denote the three parameter probit model. The notations are the same as those discussed in the body of Subsection XIII.C. The model is fitted to the data by maximum likelihood estimation, as discussed in detail in Section X. Let $\hat{p}_o, \hat{\beta}_o, \hat{\beta}_1$ denote the parameter estimates and let $\hat{\sigma}(\hat{p}_o), \hat{\sigma}(\hat{\beta}_o), \hat{\sigma}(\hat{\beta}_1), R$ denote their estimated asymptotic standard errors and the asymptotic correlation matrix. These estimates are obtained directly from the outputs describing the probit model fits. Several such outputs are illustrated in Section X. The estimated asymptotic variance-covariance matrix of $\hat{p}_o, \hat{\beta}_o, \hat{\beta}_1$ is thus

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}(p_o) & o & o \\ o & \hat{\sigma}(\hat{\beta}_o) & o \\ o & o & \hat{\sigma}(\hat{\beta}_1) \end{pmatrix} \hat{R} \begin{pmatrix} \hat{\sigma}(p_o) & o & o \\ o & \hat{\sigma}(\hat{\beta}_o) & o \\ o & o & \hat{\sigma}(\hat{\beta}_1) \end{pmatrix}$$

Let Z_o, Z_i denote the Z-values corresponding to the concentrations at the control group and at treatment group i respectively. The difference between the mortality rates at these two groups is

$$f(p_o, \beta_o, \beta_1; Z_o, Z_i) \equiv (1-p_o)[\Phi(\beta_o + \beta_1(Z_i - m)) - \Phi(\beta_o + \beta_1(Z_o - m))]$$

This difference is estimated by substituting $\hat{p}_o, \hat{\beta}_o, \hat{\beta}_1$ in the above expression.

To construct a confidence interval on the difference we must calculate the asymptotic standard error of the estimated difference. This is done by the delta method. Namely

$$\partial f/\partial p_o = -[\Phi(\beta_o + \beta_1(Z_i - m)) - \Phi(\beta_o + \beta_1(Z_o - m))]$$

$$\partial f/\partial \beta_o = (1-p_o)[\phi(\beta_o + \beta_1(Z_i - m)) - \phi(\beta_o + \beta_1(Z_o - m))]$$

$$\partial f/\partial \beta_1 = (1-p_o)[(Z_i - m)\phi(\beta_o + \beta_1(Z_i - m)) - (Z_o - m)\phi(\beta_o + \beta_1(Z_o - m))]$$

$$\hat{\text{Var}}(\hat{f}) \doteq (\partial\hat{f}/\partial p_o, \partial f/\partial \beta_o, \partial\hat{f}/\partial\beta_1)\hat{\Sigma}(\partial\hat{f}/\partial p_o, \partial\hat{f}/\partial\beta_o, \partial\hat{f}/\partial\beta_1)'$$

$$\hat{\text{Std err}}(\hat{f}) = [\hat{\text{Var}}(\hat{f})]^{1/2}$$

In the above expressions, $\hat{f}, \partial\hat{f}/\partial p_o, \partial\hat{f}/\partial\beta_o, \partial\hat{f}/\partial\beta_1$ are obtained by substituting $\hat{p}_o, \hat{\beta}_o, \hat{\beta}_1$ for p_o, β_o, β_1. The expression $\phi(\cdot)$ denotes the standard normal probability density function. If logarithmic concentration is used and if concentration is 0 at the control group then $\Phi(\beta_o + \beta_1(Z_o - m))$, $\phi(\beta_o + \beta_1(Z_o - m))$, and $(Z_o - m)\phi(\beta_o + \beta_1(Z_o - m))$ are set identically to 0.

A $1-\alpha$ level two sided confidence interval on the difference is constructed as

$$(\hat{f}-\xi_{1-\alpha/2}\hat{\text{Std err}}\ (\hat{f}),\hat{f}+\xi_{1-\alpha/2}\hat{\text{Std err}}(\hat{f}))$$

where $\xi_{1-\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution.

Simultaneity can be adjusted for by Bonferroni's method, but has not been done so for these intervals.

DISTRIBUTION LIST

25 copies	Commander ATTN: SGRD-UBG US Army Medical Bioengineering Research and Development Laboratory Fort Detrick, Frederick MD 21701-5010
4 copies	Commander US Army Medical Research and Development Command ATTN: SGRD-RMS Fort Detrick, Frederick, Maryland 21701-5012
1 copy	Defense Technical Information Center (DTIC) ATTN: DTIC-DDAC Cameron Station Alexandria, VA 22304-6145
1 copy	Dean School of Medicine Uniformed Services University of the Health Sciences 4301 Jones Bridge Road Bethesda, MD 20814-4799
1 copy	Commandant Academy of Health Sciences, US Army ATTN: AHS-CDM Fort Sam Houston, TX 78234-6100
1 copy	Commander US Army Medical Bioengineering Research and Development Laboratory ATTN: SGRD-UBD-A/Librarian Fort Detrick, Frederick, MD 21701-5010

END

FILMED

5-85

DTIC