AD NUMBER

AD861067

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; SEP 1969. Other requests shall be referred to Department of the Army, Fort Detrick, Attn: Technical Releases Branch/TID, Frederick, MD 21701.

AUTHORITY

SMUFD ltr, 25 May 1970

AD

TECHNICAL MEMORANDUM 181

DATA ANALYSIS WITH TRANSFORMATIONS

H. Gill Hilton

SEPTEMBER 1969

DEPARTMENT OF THE ARMY
Fort Detrick
Frederick, Maryland

DDC AVAILABILITY NOTICES

Qualified requesters may obtain copies of this publication from DDC.

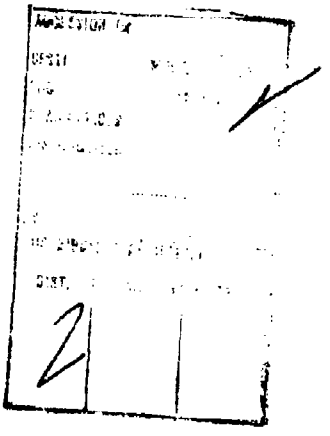
Foreign announcement and dissemination of this publication by DDC is not authorized.

Release or announcement to the public is not authorized.

DISPOSITION INSTRUCTIONS

Destroy this publication when it is no longer needed. Do not return it to the originator.

The findings in this publication are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.



DEPARTMENT OF THE ARMY
Fort Detrick
Frederick, Maryland 21701

TECHNICAL MEMORANDUM 181

DATA ANALYSIS WITH TRANSFORMATIONS

H. Gill Hilton

Biomathematics Division
ANALYTICAL SCIENCES DIRECTORATE

Project 1B562602A0D1

September 1969

FOREWORD

This memorandum was originally written as Biomathematics Division Analysis 9163 dated June 1969.

ABSTRACT

The lambda transformation of Box and Cox was applied to 56 sets of data from five areas of biological research to determine the optimum transformation for a given type of data. Data from one area of research (mask) were improved by a transformation of $\lambda_1 = 0$. This corresponds to a log transform and has been applied routinely to such data. Data from other areas of research were less affected by the transform. For these non-mask data, significance of main effects was not changed, interactions were generally unaffected, and variance homogeneity was achieved in only two of the six possible cases where the lambda transform was compared with no transform.

The study corroborates the analysis that has been performed regularly on data from one area of research but indicates that an analysis on untransformed data would generally be as meaningful for the other four areas of research examined. Where it was helpful, the main influence of the transform was in stabilizing variance. For the small (eight) number of cases examined, no real improvement in additivity was noted.

I. INTRODUCTION*

The purpose of scientific experimentation is to obtain information; many scientific data are evaluated statistically to aid in maximizing or interpreting this information. The proper use of any statistical method depends upon how closely certain necessary assumptions are satisfied by the data in question. The more commonly used statistical methods have restrictive assumptions and are usually termed "parametric" methods. Because it is generally impossible for a researcher to generate data that exactly satisfy even broad assumptions, the use of any statistical method is an approximation whose effectiveness directly correlates with how closely the data meet the necessary assumptions.

One of the statistical methods commonly used during the past few decades is analysis of variance. The proper application of analysis of variance, including tests of significance, is based upon several assumptions whose validity is rarely tested. These assumptions, in order of their probable importance, are: (i) the error variance is homogeneous; (ii) the effects are additive; and (iii) the observations are normally distributed.

The normality assumption is of little practical concern, due to the central limit theorem. The additivity assumption is not important if one places interaction terms in the model and such terms themselves are additive. The importance of the homogeneity assumption lies in the fact that if it is violated, improper errors may be used for certain comparisons, leading to loss of sensitivity in significance tests and inefficiency in estimating treatment effects.

Some types of data known to violate the above assumptions have been routinely subjected to a transformation prior to analysis. The most common transform at Fort Detrick has been logarithmic, which is appropriate when the variance is proportional to the square of the mean.** This transform is ofttimes successful in aiding additivity as well as stabilizing the variance. Certain percentage data have been subjected to the arc sine transformation. Transformations have thus been commonly used to improve the approximation to the necessary assumptions and to increase the amount of information obtainable from a given set of data. Box and Cox*** proposed a parametric transformation from y to $y^{(\lambda)}$ where

---

* This report should not be used as a literature citation in material to be published in the open literature.

** Eisenhart, Churchill. 1947. The assumptions underlying the analysis of variance. Biometrics 3:1-22.

*** Box, G.E.P.; Cox, D.R. 1964. An analysis of transformations. J. Roy. Statist. Soc. Ser. B 26:211-252.

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases} \tag{1}$$

and

$$y^{(\lambda)} = \begin{cases} \dfrac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \log (y + \lambda_2) & \lambda_1 = 0 \end{cases} \tag{2}$$

Under the assumption that for some unknown λ the transformed observations $y^{(\lambda)}$ satisfied the full normal theory analysis of variance assumptions, the maximum likelihood estimate for λ was found. The above-mentioned authors chose to express their results in terms of the normalized transformation $Z^{(\lambda)}$ where

$$Z^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda - 1}} \tag{3}$$

and where $\dot{y}$ is the geometric mean of the observations or

$$Z^{(\lambda)} = \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1 \{gm(y + \lambda_2)\}^{\lambda_1 - 1}} \tag{4}$$

where $gm(y + \lambda_2)$ is the geometric mean of $(y + \lambda_2)$.

If $\lambda_1 = 0$, $Z^{(\lambda)} = \{gm(y + \lambda_2)\} \log (y + \lambda_2)$ (5)

The maximized log likelihood, $L_{max}(\lambda)$, was equal to

$$-\frac{1}{2} n \log \hat{\sigma}^2(\lambda; Z) \tag{6}$$

where

$$\hat{\sigma}^2(\lambda; Z) = \frac{S(\lambda; Z)}{n} \tag{7}$$

where $S(\lambda; Z)$ is the residual sum of squares of $Z^{(\lambda)}$. The maximum likelihood estimate of λ is thus that value of λ that minimizes $S(\lambda; Z)$.

Procedures were furnished by Box and Cox that enable one to determine the relative contribution to the estimate of λ from normality, homogeneity of variance, and additivity. That is, one can estimate λ so as to most nearly achieve normality, to most nearly achieve normality and homogeneity, and to most nearly achieve normality and homogeneity and additivity. The three separate estimates of λ are not always the same, i.e., it is not always possible to find a single transformation that will simultaneously achieve normality, homogeneity, and additivity. All that is necessary to estimate λ is some appropriate estimate of experimental error, but if one wishes to delineate the contributions to λ of the three criteria, some specific inputs are necessary. In an analysis of variance or a multiple regression context one can speak of within-cell variance, second-order effects, and third- or higher-order effects. Table 1 shows the different estimates of error as they relate to the three criteria. It is thus necessary to have data from an experimental design having within-cell replication and effects of at least third order to separate the influence of normality, homogeneity, and additivity on the estimate of λ.

TABLE 1. ESTIMATES OF ERROR ASSOCIATED WITH THE THREE OPTIMUM MODEL CHARACTERISTICS

Criteria[a]	Quantities in Error Estimate
N	Within cells
H, N	Within cells and effects of third order and higher
A, H, N	Everything except first-order effects

a. N, normality; H, homogeneity; A, additivity.

One should keep in mind the comments of Box and Cox when using the technique: ". . . the method developed below for finding a transformation is useful as a guide, but is, of course, not to be followed blindly," and one should "tentatively entertain the basis for analysis," and maintain an attitude of "sceptical optimism." With the theory thus worked out it seemed appropriate to utilize the Box and Cox approach in a rather thorough look at, and comparison with, the current analyses of data from Fort Detrick investigations. Data from several different types of research, including decay rates, specific activities, plant yields, per cent penetrations in mask studies, and blood counts, were examined in an attempt to find what, if any, the optimum transformation should be for each kind of experimental data.

II. METHODS

The methods in the paper by Box and Cox were developed for fixed effects* analysis of variance models. Certain liberties have been taken with the method for use on mixed models. Essentially, various treatment x random element interactions were assumed equal so they could be pooled into a common error term that was then used as the error for testing all main effects. This is no different from what is commonly done in split plot analyses where several treatment x block interactions are pooled to form the estimate of split plot error.

The computations performed during this study were facilitated by a CDC 3150 electronic digital computer. Three different computer programs were used for various parts of the study, two adaptations of existing programs** and the third*** written expressly for the study. The programs are BOXNCOX, REGBXNCX, and MORBXNCX.

BOXNCOX has evolved into the major system and can easily be used on a production run basis. It finds the value of λ_1 that minimizes $S(\lambda; Z)$ in equation (7) and then allows performance of two distinct analyses of variance. One analysis will be on data that are transformed according to the optimum λ_1. The second analysis can be on the raw data or on the raw data with a log or arc sine transform. The experimenter is thus able to compare the optimum λ analysis with a standard analysis.

For nonorthogonal data (but assuming fixed effects), the program REGBXNCX was modified from an existing multiple regression program. It finds the optimum value of λ that minimizes $S(\lambda; Z)$ in equation (7) and then allows an analysis on the data thus transformed and also for any other transform desired for purposes of comparison.

For data arising from an experimental design involving two or more crossed factors and true within-cell replication, it is possible to get some idea as to the relative contributions of normality, model simplicity, and variance homogeneity to the estimate of λ. Program MORBXNCX was written to furnish part of this information when the investigator desires such a detailed breakdown.

---

* Eisenhart, Churchill. 1947. The assumptions underlying the analysis of variance. Biometrics 3:1-22.

** Dr. Roebert L. Stearman wrote one of these programs.

*** James F. Jacobs and Brucy C. Gray assisted with this program.

III. RESULTS

Experimental data from several different areas of research were examined by the Box and Cox technique. These data arose from investigations on mask efficiencies, plant response to chemicals, log source strengths and decay rates from aerosols, blood parameters, and specific activities of a biologic system.

A. PLANT STUDIES

Eight experiments involving four kinds of responses were analyzed. Table 2 shows the experiment size and response variable, and Table 3 gives the estimates of λ, F values for two-factor interactions, and treatments for both the raw and transformed data. Values of λ ranged from -0.46 to 1.23. The optimum transformation had little effect on removing additivity or improving sensitivity; i.e., the F tests for interaction and treatments were not essentially different when comparing the raw data with the transformed data. Experiment 8, involving 720 data points, had true within-cell replication, which allowed an examination of the within-cell variance as influenced by the transformation. Figure 1 shows a plot of Bartlett's variance test* versus several values of λ_1. There is no value of λ_1 that will stabilize the variance but the value that minimizes the variance heterogeneity, $\lambda_1 = 0.7$, is close to the overall optimum λ, which suggests that the main contribution to the estimate of λ comes from the homogeneity criterion.

TABLE 2. SIZE AND RESPONSE VARIABLE OF EIGHT EXPERIMENTS FROM PLANT SCIENCES

Exp. No.	No. of Data Points	Response Variable
1	58	Fresh weight of beans
2	46	Fresh weight of beans
3	75	Fresh weight of beans
4	38	Fresh weight of beans
5	20	Fresh weight of tree seedlings
6	20	Fresh weight of tree seedlings
7a,b,c,d,e	30 each	Spore count
8	720	Abscission force on bean leaves

* Bartlett, M.S. 1947. The use of transformations. Biometrics 3:39-52.

TABLE 3. LAMBDA ESTIMATES AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS FOR UNTRANSFORMED AND TRANSFORMED DATA FROM EIGHT PLANT SCIENCE EXPERIMENTS

Exp. No.	λ	F - Main Effects		F - Interactions	
		Untransformed	Transformed	Untransformed	Transformed
1	0.56	63.69[a]	48.69[a]	0.075	0.051
2	1.23	22.15[a]	23.59[a]	0.384	0.435
3	0.29	81.98[a]	88.86[a]	1.137	1.471
4	0.18	28.28[a]	24.21[a]	-[b]	-
5	-0.44	5.204	5.508	-	-
6	-0.46	2.149	2.207	0.033	0.008
7 a	0.78	11.31[a]	10.96[a]	4.923[a]	4.863[a]
b	0.47	13.47[a]	13.58[a]	42.04[a]	47.061[a]
c	0.83	53.04[a]	63.34[a]	17.31[a]	15.363[a]
d	0.75	30.40[a]	36.67[a]	3.746[a]	2.440
e	-0.12	13.79[a]	44.32[a]	2.363	1.246
8	0.76	38.68[a]	39.61[a]	4.512[a]	4.704[a]

a. Indicates significance at $P \leq 0.05$.
b. Hyphen indicates no interaction could be estimated.

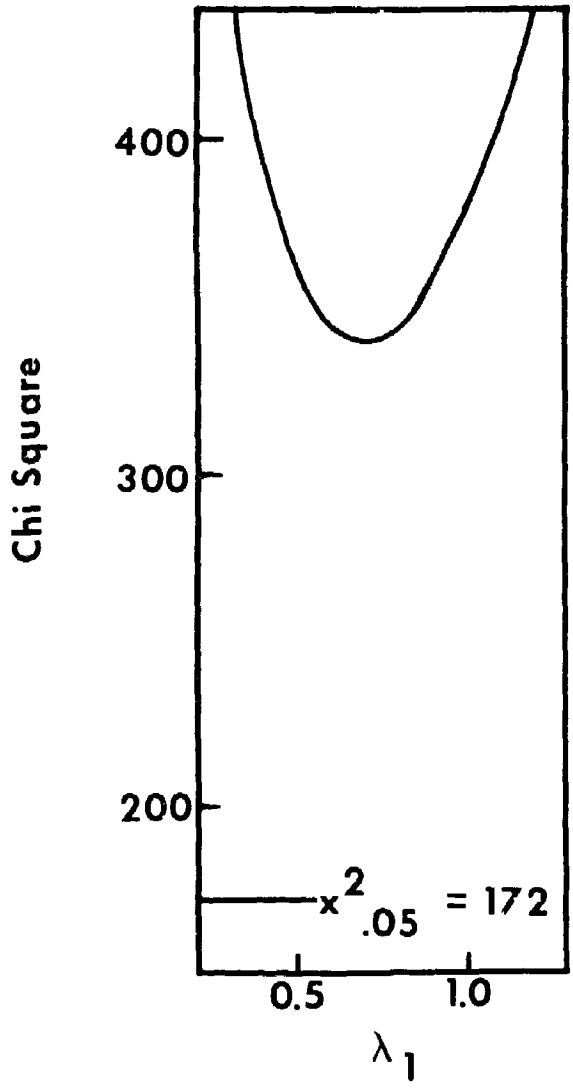


FIGURE 1. Homogeneity of Variance Criterion Versus λ_1 for Bean Data from Experiment 8. Critical value of chi square is shown by horizontal line.

The relative contributions to λ of normality, homogeneity, and additivity are shown in Figure 2 for bean data from experiment 8. About all one can say about normality is that apparently a rather wide choice of λ will give similar log likelihoods. Some data (Fig. 3) give a much broader curve for N when plotting log likelihood versus λ. When one adds the restriction of homogeneity to normality, the estimate of λ sharpens considerably, as seen in Figure 2. If the AHN curve is superimposed on the HN curve, they are practically identical, which shows that the additivity restriction adds nothing to the estimate of λ in both Figures 2 and 3.

B. MASK STUDIES

Ten different experiments of a similar type were analyzed. The response variable for each experiment was per cent penetration of an aerosol into a mask. The experiment size and estimates of λ_1 are listed in Table 4. If a weighted average of these 10 λ's is computed using error degrees of freedom as weights, the mean λ is equal to 0.01, not essentially different from zero. The overall F values for significance of the pooled main effects and the F values for significance of the two-factor interactions for both the transformed and raw data are shown in Table 4. Significance of main effects was generally enhanced, with significant effects present in only two of the experiments prior to transform but in six of the experiments after transform. The two-factor interactions were nonsignificant both before and after transformation. This lack of interaction is probably due to the very careful conditions under which the experiments are conducted.

C. SPECIFIC ACTIVITY

Results from 10 experiments are listed in Table 5. Estimates of λ, while somewhat variable, average 0.85, not essentially different from 1.0. The F tests for significance of pooled two-factor interactions and for pooled main-effect treatments did not change materially when comparing the transformed results with those from the raw data. Values of F for testing treatments and interactions varied from the raw analysis to the transformed analysis, but the only meaningful change was with one experiment in which the treatment effects became nonsignificant after transformation whereas they had been significant on the raw data.

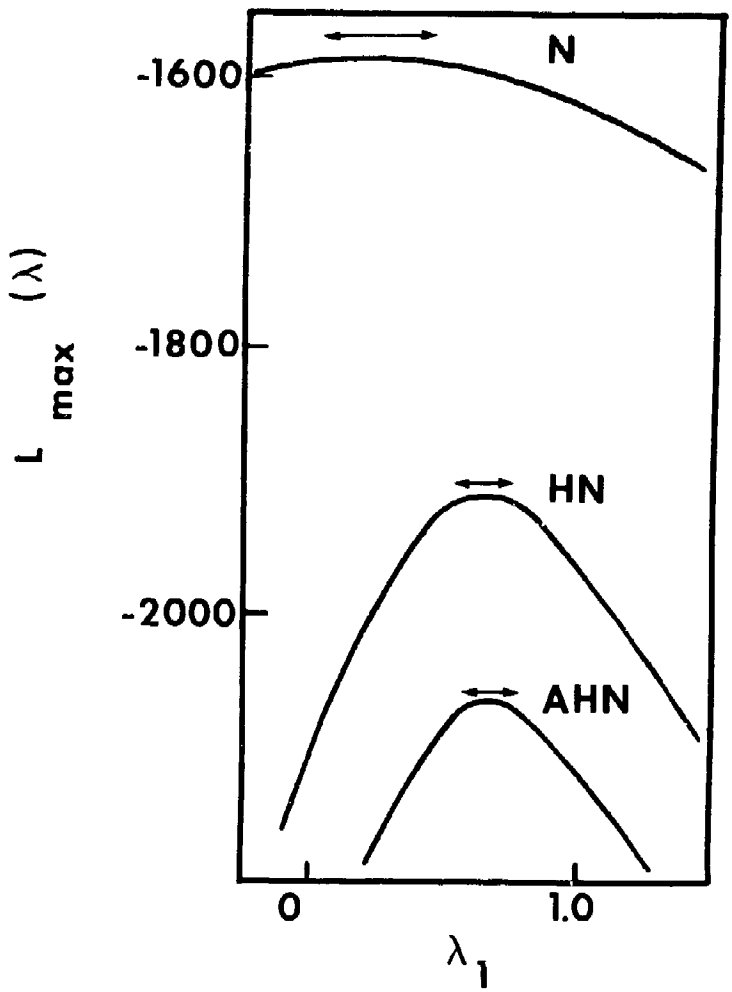


FIGURE 2. Log Maximum Likelihood Versus λ_1 as Influenced by Normality, N; Homogeneity, H; and Additivity, A, for Bean Data from Experiment 8. Arrows indicate approximate 95% confidence limits on λ_1.

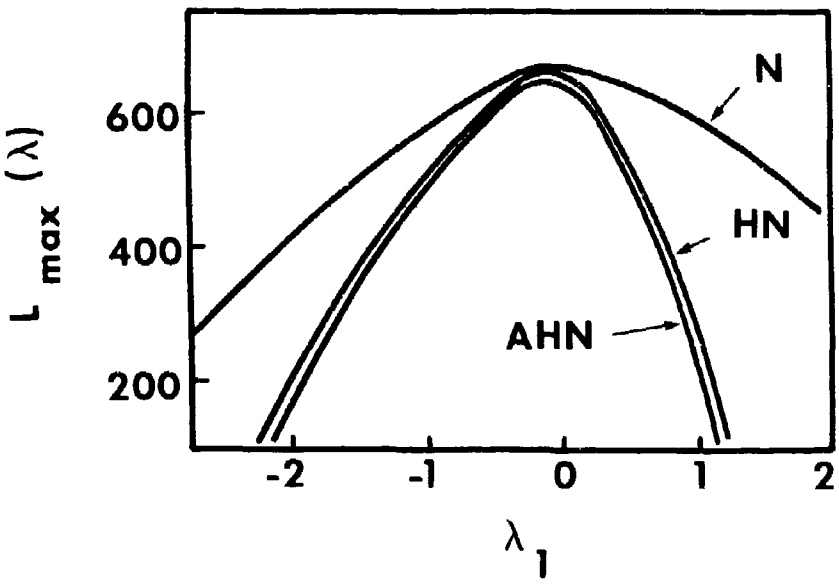


FIGURE 3. Log Maximum Likelihood Versus λ_1 as Influenced by Normality, N; Homogeneity, H; and Additivity, A, for Mask Data from Experiment 10.

TABLE 4. NUMBER OF OBSERVATIONS, λ ESTIMATES, AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS FOR TRANSFORMED AND RAW DATA FROM 10 MASK EXPERIMENTS

Exp. No.	No. of Observations	λ	F - Main Effects		F - Interactions	
			Untransformed	Transformed	Untransformed	Transformed
1	81	-0.55	2.271	3.125	1.302	1.169
2	72	0.34	13.72[a]	18.52[a]	1.508	1.047
3	72	-0.09	1.037	2.640[a]	0.851	1.135
4	72	0.17	3.577[a]	5.417[a]	1.378	0.891
5	72	-0.17	0.827	0.697	0.357	0.598
6	72	0.33	0.856	1.133	0.838	0.700
7	72	0.19	1.212	2.003	1.169	0.923
8	72	0.36	1.926	2.140[a]	0.832	0.782
9	72	-0.15	1.556	3.991[a]	0.783	0.690
10	90	-0.22	1.389	2.704[a]	0.755	1.223
Mean		0.01[b]				

a. Indicates significance at $P \leq 0.05$.
b. Weighted mean with error degrees of freedom as weights.

TABLE 5. NUMBER OF OBSERVATIONS, λ ESTIMATES, AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS FOR UNTRANSFORMED AND TRANSFORMED DATA FROM 10 SPECIFIC ACTIVITY EXPERIMENTS

Exp. No.	No. of Observations	λ	F - Main Effects		F - Interactions	
			Untransformed	Transformed	Untransformed	Transformed
1	64	1.21	0.920	1.810	0.625	0.570
2	64	0.32	3.611[a]	3.726[a]	0.964	0.951
3	64	0.18	2.755[a]	2.126	2.306[a]	3.876[a]
4	64	1.48	1.961	1.848	0.335	0.290
5	64	0.79	2.709[a]	2.815[a]	1.280	1.409
6	64	1.06	3.461[a]	3.299[a]	0.956	0.923
7	64	1.18	4.031[a]	3.472[a]	2.411[a]	4.842[a]
8	64	0.59	5.467[a]	5.363[a]	1.280	1.278
9	64	1.09	4.306[a]	4.312[a]	0.976	0.968
10	64	0.57	7.881[a]	8.128[a]	1.285	1.448
Mean		0.85				

a. Indicates significance at $P \leq 0.05$.

D. BLOOD PARAMETERS

Twelve different blood parameters were measured on animal blood samples collected daily for 6 days. Because some of the responses were negative, an additive constant was necessary to allow an estimate of λ_1. These constants, labeled λ_2, along with the experiment size and response variable are listed in Table 6. A separate analysis was performed on each parameter (Table 7). Estimates of λ ranged from -3.01 to 2.02. The transformation had little effect on the results since the treatment differences were significant in only one of the 12 cases where they were not significant prior to the transform. Only one of the 12 cases had significant two-factor interactions, but this was removed on the transformed metric.

TABLE 6. SIZE AND RESPONSE VARIABLE OF A BLOOD PARAMETER STUDY

Exp. No.	No. of Observations	λ_2	Response Variable
1	57	1.0	Basophil
2	57	4.0	Monocytes
3	57	44.0	Lymphocytes
4	57	1.0	Metamyelocytes
5	57	4.0	Eosinophil
6	57	3.0	Non-segmenter cells
7	57	37.0	Segmenter cells
8	59	6,513.0	Total white blood cells
9	59	2.2	Per cent reticulocytes
10	59	6.0	Hemoglobin
11	59	16.0	Packed cell volume
12	59	3.83	Red blood cells

TABLE 7. LAMBDA ESTIMATES AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS FOR UNTRANSFORMED AND TRANSFORMED DATA FROM A BLOOD PARAMETER STUDY

Exp. No.	λ_1	F - Main Effects		F - Interactions	
		Untransformed	Transformed	Untransformed	Transformed
1	-2.96	1.230	6.142[a]	1.325	2.096
2	0.74	2.533[a]	2.559[a]	0.733	0.678
3	1.14	5.480[a]	5.017[a]	0.320	0.316
4	-3.01	1.591	1.069	2.458[a]	1.751
5	0.83	2.811[a]	2.978[a]	1.314	1.192
6	-0.71	2.654[a]	7.294[a]	1.829	1.408
7	1.44	4.217[a]	3.614[a]	0.516	0.467
8	1.31	5.861[a]	5.108[a]	1.236	1.143
9	0.67	2.897[a]	2.919[a]	0.309	0.379
10	1.81	3.838[a]	4.111[a]	1.251	1.247
11	1.78	2.477[a]	2.944[a]	1.578	1.451
12	2.02	4.838[a]	5.290[a]	0.901	0.796

a. Indicates significance at $P \leq 0.05$.

E. AEROSOL CHAMBER EXPERIMENTS

Results from six aerosol studies are listed in Tables 8 and 9. The response variables come from a regression line associated with a particular treatment in the experiment. The slope of the line is associated with the decay rate of the aerosol and the intercept with the source strength. Estimates of λ for decay rates (Table 8) are reasonably stable and only vary from 1.16 to 1.69 with an average of about 1.4. However, no effect on the interpretation of the experimental results is noted due to the transform either in terms of main effects or interactions.

Log source strength data gave rise to more variable estimates of λ with values ranging from -1.42 to 3.09. No effect on interpretation was noted due to transform, although, as shown in Table 9 for experiment 3, the transform did yield an F value approximately three times that of the untransformed data.

TABLE 8. DECAY RATES FROM SOME CHAMBER AEROSOL EXPERIMENTS: NUMBER OF DATA POINTS, λ ESTIMATES, AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS

Exp. No.	No. of Observations	λ	F - Main Effects		F - Interactions	
			Untransformed	Transformed	Untransformed	Transformed
1	36	1.26	1.050	0.973	-[a]	-
2	36	1.16	1.544	1.482	-	-
3	40	1.48	3.604[b]	2.973[b]	-	-
4	25	1.45	1.710	1.766	-	-
5	32	1.38	2.090	2.229	1.655	1.755
6	16	1.69	1.008	0.901	0.747	0.845

a. Hyphen indicates no interaction could be estimated.
b. Indicates significance at $P \leq 0.05$.

TABLE 9. LOG SOURCE STRENGTHS FROM SOME CHAMBER AEROSOL EXPERIMENTS: NUMBER OF DATA POINTS, λ ESTIMATES, AND F VALUES FOR MAIN EFFECTS AND TWO-FACTOR INTERACTIONS

Exp. No.	No. of Observations	λ	F - Main Effects		F - Interactions	
			Untransformed	Transformed	Untransformed	Transformed
1	36	0.49	3.302[a]	2.828[a]	-[b]	-
2	36	2.27	8.325[a]	8.939[a]	-	-
3	40	-1.42	12.59[a]	34.50[a]	-	-
4	25	2.84	2.514[a]	2.520[a]	-	-
5	32	1.79	2.499	2.276	1.452	1.299
6	16	3.09	1.524	1.196	1.906	1.325

a. Indicates significance at $P \leq 0.05$.
b. Hyphen indicates no interaction could be estimated.

IV. DISCUSSION

The results from 56 sets of data for λ from the 46 separate experiments from five different areas of research gave somewhat similar outcomes. For point of reference, recall that a value of $\lambda_1 = 1$ is equivalent to no transform or to an analysis on the raw data itself. Every analysis performed had a λ_1 different from 1, but with one broad exception there was little effect on the results by following the λ transform. This exception was associated with the mask studies where the suggested transform was approximately $\lambda_1 = 0$, which is equivalent to the log transform. Such data have been analyzed routinely on a log transform basis, so the Box and Cox study corroborates accepted practice very nicely. However, estimates of λ_1 ranged from -0.55 to 0.36 for the individual mask studies evaluated, suggesting that a particular set of data is more efficiently analyzed by its own λ_1, although on the average an overall common transform is apparently satisfactory for these data.

While estimates of λ_1 were always different from 1 for the other types of data, following the transformed analysis compared with the untransformed analysis gave no apparent gain. Generally speaking, if an effect was significant, it was significant in both transformed and untransformed analyses. There were isolated exceptions, but not enough that couldn't be explained as being due to the equivalent of a type I error. Of the 46 data sets other than the mask studies, 33 had significant main effects with both the transformed and untransformed analyses. In only one case did the transform lead to a significant result that was not significant in the untransformed data, and in one case the converse was true. Thus, there were 31 of the 46 non-mask studies where no added influence on significance of main effects due to transformation was noted. Since the 13 data sets having nonsignificant main effect were not altered due to the transform, one could state that in 44 of the 46 cases studied, the Box and Cox analysis did not essentially affect the interpretation.

One goal of the Box and Cox transform was to simplify the model, i.e., to eliminate interaction constants from the model. Of the 56 separate data sets, 46 were such that a test for two-factor interactions was possible. Only eight of these 46 cases had significant two-factor interactions on untransformed data, and six of the eight were still significant after the transform. Most of the significant interactions were associated with the crops experiments. The other four experimental areas apparently have such careful control of their experimental conditions or are experimenting over such a narrow range of treatments that interactions did not generally appear. Many of the crops experiments were exploratory in nature with accompanying wide ranges on factor levels that naturally lead to interactions in most biologic systems.

Probably the most desired quality of any transformation is its ability to stabilize the variance. In order to examine the variance-stabilizing ability of the Box and Cox λ transform, it is necessary to have data that come from a design with true within-cell replication. Only 11 of the 56 data sets studied had this characteristic, but 17 others were such that a fairly simple assumption made it possible to act as if cell replication were present. These 28 cases were composed of nine mask studies, seven crops studies, and 12 blood studies. The 28 cases were examined in detail to determine the effect of transform on variance stability. In 13 of the 28 the variance was stable with no transformation, but with the λ transform 23 of the 28 studies had stable variances. None of the mask studies had stable variance without the transform, whereas with the crops studies the variance was stable both with and without transform. Of the five studies in which it was impossible to achieve a stable variance, two were blood studies, two were crops studies, and one was a mask study.

The results from the studies on variance stability indicate that the λ transform has little effect on this characteristic of data for the crops and blood studies. In contrast, the λ transform most markedly affected variance stability for the mask data. In only one of the nine mask studies where variance stability could be examined did the λ transform fail to stabilize the variance. It appears that the primary value of the transform for mask data is in achieving stability of variance.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Department of the Army Fort Detrick, Frederick, Maryland, 21701	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
DATA ANALYSIS WITH TRANSFORMATIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
H. Gill Hilton

6. REPORT DATE September 1969	7a. TOTAL NO. OF PAGES 25	7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. b. PROJECT NO. 1B562602A0D1 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Memorandum 181 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT
Qualified requesters may obtain copies of this publication from DDC.
Foreign announcement and dissemination of this publication by DDC is not authorized.
Release or announcement to the public is not authorized.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Department of the Army Fort Detrick, Frederick, Maryland, 21701

13. ABSTRACT

The lambda transformation of Box and Cox was applied to 56 sets of data from five areas of biological research to determine the optimum transformation for a given type of data. Data from one area of research (mask) were improved by a transformation of $\lambda_1 = 0$. This corresponds to a log transform and has been applied routinely to such data. Data from other areas of research were less affected by the transform. For these non-mask data, significance of main effects was not changed, interactions were generally unaffected, and variance homogeneity was achieved in only two of the six possible cases where the lambda transform was compared with no transform.

The study corroborates the analysis that has been performed regularly on data from one area of research but indicates that an analysis on untransformed data would generally be as meaningful for the other four areas of research examined. Where it was helpful, the main influence of the transform was in stabilizing variance. For the small (eight) number of cases examined no real improvement in additivity was noted.

14. Key Words

Transformation
Homogeneity of variance
Analysis of variance
Data analysis

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification