AD669899

NWL Technical Report
TR-2108

THE METHOD AND USE OF NOVACOM, A PROGRAM FOR "NON-ORTHOGONAL" ANALYSIS OF VARIANCE AND COVARIANCE

By

Klaus Abt

19 April 1968

Best
Available
Copy

ABSTRACT

The report contains the description of a program ("NOVACOM") for the solution of problems in the area of analyzing data based on the general linear statistical model. While the detailed program documentation is given elsewhere, the present publication deals with the statistical method, the logical flow, and the use and application of NOVACOM in multiple linear regression and ("non-orthogonal") analysis of variance and covariance for crossed classifications with incomplete and unbalanced data. The method of NOVACOM is basically a backward ranking procedure applied to individual and/or groups of independent variables (concomitant independent variables and/or ANOVA effects, respectively). The result of the ranking is a model ("significant model") which contains only significant concomitant independent variables and/or ANOVA effects. The method and use of the program is illustrated by examples of the statistical analysis of bodies of incomplete experimental data.

FOREWORD

The work covered by this report was done in the Mathematical Statistics Branch of the Operations Research Division, Computation and Analysis Laboratory, under Foundational Research Project No. 29Y, "Computer Programs for Statistical Analyses."

The report contains the description of the method and use of the computer program NOVACOM, which performs analysis of variance and covariance for unbalanced data classifications with missing values, i.e., for situations which are often met in the analysis of Naval ordnance experimentation and test data.

NOVACOM was coded from notes (similar in content to some parts of the present report) by Mr. T. Herring of the Programming Division, Computation and Analysis Laboratory. Mr. Herring, who contributed significantly also to the general methodological concept of NOVACOM, is the author of the program documentation of NOVACOM. ("A Programming Guide to NOVACOM", NWL Technical Memorandum, in preparation.)

Many ideas for the concept of NOVACOM were contributed by Messrs. C. Bates, G. Gemmill and R. Shade of the Mathematical Statistics Branch. Mr. A. R. DiDonato and Dr. M. P. Jarnagin of the Mathematics Research Group, Computation and Analysis Laboratory, developed the method of the subroutine ISUBX for the computation of the incomplete beta function ratio contained in NOVACOM. This method is documented in NWL Report No. 1949, revised October 1966.

The author wishes to thank Dr. Sidney Addelman and Mr. James Merrill of the Research Triangle Institute, Durham, North Carolina, for their valuable comments on the interpretation of the NOVACOM results.

The report was typed by Miss Judy D. Merryman.

The work on this report was completed on 28 July 1967.

APPROVED FOR RELEASE:

Bernard Smith

BERNARD SMITH
Technical Director

TABLE OF CONTENTS

	Page
ABSTRACT	ii
FOREWORD	iii
1. INTRODUCTION	1
2. METHOD OF NOVACOM	3
2.1 General Outline of the Method	3
2.1.1 The Model	3
2.1.2 The Backward Ranking Method	9
2.1.3 Accuracy Checks on Matrix Inversions	18
2.1.4 Printout and Comprehensive Analyses	19
2.2 Automatic Generation and Controls	20
2.2.1 Generation and Admissibility of CIVs	20
2.2.2 Generation of DIVs and Admissibility of Factorial Effects	23
2.2.3 Generation of the Design Matrix	34
2.3 The Ranking Subroutines of NOVACOM	37
2.3.1 COMO	37
2.3.2 FEMO	41
2.4 Computational Flow	50
2.4.1 Flowcharts	50
2.4.2 Comments on Flowcharts	66
3. USE OF NOVACOM	72
3.1 Input Preparation	72
3.1.1 Control Cards and Data Cards	72
3.1.2 Model Generation Options	80
3.1.3 Ranking Options	83
3.2 Running Time Formula	87
3.3 Interpretation of Results	88
3.3.1 The Final Comprehensive Analysis (FCA)	88
3.3.2 The Final FCA	91
3.3.3 Additional Analyses of Variance (ANVAs)	92
3.4 Examples of Application	93
3.4.1 Example 1	95
3.4.2 Example 2	117
3.4.3 Example 3	124
3.4.4 Example 4	134
3.4.5 Example 5	140
3.4.6 Example 6	147
3.4.7 Example 7	171
4. GLOSSARY OF TERMS USED IN THE REPORT	176
5. REFERENCES	181
APPENDICES	
A. Method of Fitting Constants for Non-Orthogonal Layouts with Interactions and Empty Cells	
B. Distribution	

1. INTRODUCTION

The concept of the program NOVACOM ("Non-Orthogonal VAriance and COvariance analysis by Multiple Regression techniques"), as described in this report, is based on the multiple regression approach to analysis of variance (see, for example, Brownlee [1960]). The underlying method of NOVACOM was developed in such a manner that a wide variety of problems in the area of analyzing data based on the general linear statistical model can be solved. The possible applications of NOVACOM in this area are: multiple linear regression including polynomial regression, ("orthogonal") analysis of variance and covariance for crossed classifications with balanced and complete or incomplete data, and ("non-orthogonal") analysis of variance and covariance for crossed classifications with incomplete and unbalanced data when possibly some of the ANOVA effects are confounded.

While regression and "orthogonal" ANOVA may be considered as bonus areas of application, it was the third area (of "non-orthogonal" analysis of variance and covariance for crossed classifications) for which NOVACOM was developed. Most of the theory for the method of the program is described in the present report. A more detailed outline of the theory is contained in another paper by the author (Abt [1967]).

The method of NOVACOM is basically a backward ranking procedure applied to individual independent variables and/or groups of independent variables, where the groups represent analysis of variance effects in the general linear model. The ranking is done by order of prediction power for the dependent variable (response variable), where the so-called "non-significance" serves as criterion for establishing the ranking. In ranking the independent variables of analysis of covariance models, the program makes an internal decision (based upon a significance level α specified by the user) whether to include the covariates for the analysis of variance part of the ranking. There is some restriction in the ranking of ANOVA effects in that at a given step of the ranking only those effects are admissible for ranking whose associated sums of squares are independent from the restrictions chosen to make the linear model a model of full rank. (The admissibility is internally determined by the program.) The ranking procedure leads to a significant model which contains those effects (and covariates, if any) which are significant at a level specified by the user, plus those effects, if any, which did not become admissible in the ranking procedure. Accordingly, NOVACOM may be considered as a screening tool for significant factorial effects in (crossed) data classifications with possibly highly incomplete and unbalanced data. A special additional feature allows for the screening for "the most probable significant model" when there are confounded ANOVA effects.

The model in NOVACOM may include up to 139 independent variables. The limitation on the number of covariates is determined by the number of independent variables representing ANOVA effects. The factors may have qualitative and/or quantitative levels. In one given problem, up to four

different dependent variables may be analyzed, however, each one in a univariate manner and for the same values of the independent variables.

Some parts of the method of NOVACOM were taken from the program DA-MRCA for multiple linear regression (Abt et al. [1966]). Accordingly, the reader is often referred to this documentation of DA-MRCA which is listed as "Reference 2" in Section 4 of the present report. While the user of DA-MRCA was also able to perform, though in a somewhat cumbersome way, "non-orthogonal" analysis of variance (and covariance), he had no possibility of arriving at a significant model by ranking methods, and he had to do the generation of the model and the design matrix mostly by hand-input. In the method of NOVACOM, considerable emphasis is put on the automatic generation of the model and the design matrix.

Because of considerations regarding its size, the present report does not contain any documentation of the programming of NOVACOM. The FORTRAN IV documentation, as well as general programming notes are contained in a report by T. Herring [1967] who also programmed NOVACOM. In accordance with its title, the present report is restricted to the description of the method and use of NOVACOM. In addition, the report contains a number of numerical illustrations of the program's possible applications. Therefore, the report may serve as a manual for the user of the program, and for this purpose, the report also contains all the necessary information for operating the program and for interpreting the results. This necessitated some overlap with the contents of the afore-mentioned report by T. Herring; for example, both reports contain the description of the control and data cards. The two reports, each being self-contained, can be defined as the statistician's guide (the present report) and the programmer's guide to NOVACOM (the report by T. Herring).

2. METHOD OF NOVACOM

2.1 General Outline of the Method

2.1.1 The Model

The method of NOVACOM is based on the general linear statistical model,

$$y = \beta_0 + \sum_{\nu=1}^{N} \beta_\nu x_\nu + e, \qquad (2\text{-}1)$$

where,

y = "dependent" (random) variable

x_ν = "independent" (non-random) variables, $\nu = 1,\ldots,N$

β_ν = regression coefficients, $\nu = 1,\ldots,N$

β_0 = general constant

e = "residual", or "error" term: a random variable with expectation zero and variance σ^2, usually assumed to be independently normally distributed.

More specifically, the model (2-1) in NOVACOM is of the form

$$y = \beta_0 + \sum_{\nu=1}^{N-T} \beta_\nu x_\nu + \sum_{\nu=N-T+1}^{N} \beta_\nu x_\nu + e, \qquad (2\text{-}2)$$

where the first N-T independent variables represent analysis of variance effects and the last T independent variables represent concomitant variables ("covariates" if $0 < T < N$). Any one independent variable ("IV") of the first N-T will be referred to as a "Design Independent Variable" ("DIV"), and any one IV of the last T will be referred to as a "Concomitant Independent Variable" ("CIV"). Consequently, with T=N CIVs, the model (2-2) is that of multiple linear regression; with $0 < T < N$ CIVs, model (2-2) is that of analysis of covariance; and with T=0 it is the model of analysis of variance.

There are two types of CIVs: (1) those representing the original, physically observed variables (in other words, the linear terms), referred to as "Original CIVs" ("OCIVs") and (2) those CIVs representing polynomial terms, generated from the OCIVs, referred to as "Generated CIVs" ("GCIVs"). Therefore, T = [number of OCIVs] + [number of GCIVs].

The N-T DIVs (of the analysis of variance part of the model) require a more extensive discussion.

Since the application of NOVACOM is limited to crossed classification models, all ANOVA effects may be referred to as "factorial" effects (main effects, two-factor interactions, three-factor interactions, etc.). From the fact that the factorial effects are represented by groups of independent variables (DIVs) in the general linear statistical model (2-2), all factors in the analysis of variance must be considered as "fixed effects" factors. That is, "random effects" factors (the levels of which are randomly sampled from finite or infinite populations of levels) cannot be treated as such by the program NOVACOM.

First consider a model with only "qualitative" factors, i.e., factors whose levels correspond to qualitatively specified categories, such as "types of material", or "manufacturers". As an example, take the conventional ANOVA model for a two-way crossed classification with interaction (the two ways of the data classification corresponding to two factors, say "$\mathcal{A}$" and "$\mathcal{B}$"):

$$y_{\alpha\beta\rho} = m + a_\alpha + b_\beta + ab_{\alpha\beta} + e_{\alpha\beta\rho} \qquad (2\text{-}3)$$

with

$$\rho = 1,\ldots,R_{\alpha\beta}$$

$$\alpha = 1,\ldots,A$$

$$\beta = 1,\ldots,B.$$

Here $R_{\alpha\beta}$ is the number of observations (y) for the level combination, or cell, (α,β) of the two factors $\mathcal{A}$ and $\mathcal{B}$; and A and B are the numbers of levels of the two factors, respectively. (For ease of the following discussion the assumption will be made that $R_{\alpha\beta} > 0$ for all cells. This assumption, however, is not essential and all features of the "qualitative" model presently being discussed hold also for cases with empty cells, that is, for cases with some but not all $R_{\alpha\beta} = 0$.) Also in (2-3), the model constants (parameters) m, a_α, b_β, and $ab_{\alpha\beta}$ represent, respectively, a general constant, the effect of level α of factor $\mathcal{A}$, the effect of level β of factor $\mathcal{B}$, and the interaction effect of levels α and β in cell (α,β). The error term $e_{\alpha\beta\rho}$ is assumed to be normally, independently distributed with expectation zero and variance σ^2.

In order to treat the model (2-3) as a linear hypothesis model of full rank, the parameters a_α, b_β, and $ab_{\alpha\beta}$ must be subjected to linear restrictions such that the total number of degrees of freedom for

factorial effects is AB-1. The restrictions apply identically to the estimates of the parameters resulting from the solution of the system of the AB-1 normal equations. The set of restrictions used by Graybill [1961] is a very convenient choice with respect to computational simplicity as will be shown later:

$$\left.\begin{aligned} a_A &= b_B = 0 \\ ab_{A\beta} &= 0 \quad \text{for } \beta = 1,\ldots,B \\ ab_{\alpha B} &= 0 \quad \text{for } \alpha = 1,\ldots,A \end{aligned}\right\} \qquad (2\text{-}4)$$

This set of restrictions allows the remaining AB-1 parameters to be considered as contrasts with respect to the last levels of the factors. (For example, $a_1 = a_1 - a_A$.)

In order to further discuss the model (2-3), it is convenient to numerically specify A and B, the numbers of levels for factors $\mathcal{A}$ and $\mathcal{B}$, respectively. Let A=B=3, for example. Then there will be AB-1=8 parameters representing factorial effects in the example model.

In the multiple regression approach to analysis of variance (see, for example, Brownlee [1960]) each one of these parameters is considered as a "regression coefficient" of an auxiliary independent variable which takes on the value 1 when the respective effect is present and the value 0 when the effect is not present. (The auxiliary IVs will be given the symbols u_i.) In this approach, the general constant m is considered as the regression coefficient of a dummy IV, u_0, which has the constant value 1.

Applying the Graybill restrictions (2-4) and introducing the variables u_i into the model (2-3) with A=B=3, one can see that the u_i (except for u_0) represent the N-T=N-6 "DIVs" for qualitative factorial effects in the general model (2-2):

$$y_{\alpha\beta\rho} = mu_0 + a_1u_1 + a_2u_2 + b_1u_3 + b_2u_4 + ab_{11}u_5 + ab_{12}u_6 + ab_{21}u_7 + ab_{22}u_8 + e_{\alpha\beta\rho}. \qquad (2\text{-}5)$$

For cell $(\alpha,\beta) = (1,2)$, for example, the model (2-5) becomes:

$$\left.\begin{aligned} y_{12\rho} &= m\cdot 1 + a_1\cdot 1 + a_2\cdot 0 + b_1\cdot 0 + b_2\cdot 1 \\ &\quad + ab_{11}\cdot 0 + ab_{12}\cdot 1 + ab_{21}\cdot 0 + ab_{22}\cdot 0 + e_{12\rho} \\ &= m + a_1 + b_2 + ab_{12} + e_{12\rho} \ . \end{aligned}\right\} \qquad (2\text{-}6)$$

As can be seen, the values of the DIVs for interaction constants are the products of the values of the DIVs for the corresponding main effect constants. For example, $u_8 = 1 = u_1 u_4 = 1\cdot 1$; and $u_9 = 0 = u_2 u_4 = 0\cdot 1$. This "product rule" (when using restrictions of the type (2-4)) applies generally to all crossed classification models with qualitative factors. For example, in a three-way classification where factors $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ have A=2, B=3, and C=3 levels, respectively, the numerical values of the DIV attached to the interaction constant abc_{121} is $1\cdot 1\cdot 1=1$ for the R_{121} observations in cell (1,2,1) and it is 0 for the observations of all other cells. The product rule allows a simple generation of the matrix of the coefficients of the normal equations. (See Section 2.2.)

Next, consider a model with only "quantitative" factors, i.e., with factors whose levels are specified by numerical values of continuous variables, such as "temperature", or "pressure." If the two-way crossed classification is again taken as an example, the model for this case can be written:

$$y_{\alpha\beta\rho} = m u_0 + \sum_{\mu=1}^{A-1} \beta_a^{(\mu)} X_{a\alpha}^{\mu} + \sum_{\nu=1}^{B-1} \beta_b^{(\nu)} X_{b\beta}^{\nu} + \sum_{\mu=1}^{A-1}\sum_{\nu=1}^{B-1} \beta_{ab}^{(\mu,\nu)} X_{a\alpha}^{\mu} X_{b\beta}^{\nu} + e_{\alpha\beta\rho}. \qquad (2\text{-}7)$$

Here, the $X_{a\alpha} = X_a(\alpha)$ and $X_{b\beta} = X_b(\beta)$ are the numerical values of the continuous variables X_a and X_b which specify the levels of the quantitative factors $\mathcal{A}$ and $\mathcal{B}$, respectively. (Accordingly, X_a and X_b will be called "quantitative factor variables.") That is, model (2-7) is that of polynomial regression in the usual sense, and the AB-1 parameters $\beta_a^{(\mu)}$, $\beta_b^{(\nu)}$, and $\beta_{ab}^{(\mu,\nu)}$ are the regression coefficients in the usual sense. (NOTE. In "orthogonal" analysis of variance one would write the model, for this case of all factors being quantitative, in the conventional way (2-3) rather than in the way of (2-7). In the course of the analysis, one would decompose the ANOVA effects into orthogonal contrasts, for example, for factor $\mathcal{A}$, into the linear contrast, the quadratic contrast, . . ., the contrast of $(A-1)^{th}$ order. Since in "non-orthogonal" ANOVA orthogonal contrasts do not exist, the form (2-7) of the model is used here.)

As already implied in the form of the model (2-7), also in this case the DIVs representing interactions can be generated from the DIVs of the corresponding main effects by multiplication. For example, the DIV representing the $\mathcal{A}_{linear} \times \mathcal{B}_{quadratic}$ interaction,

i.e., $X_a X_b^2$, is the product of the DIVs representing the linear effect of α and the quadratic effect of β.

Finally, consider an ANOVA model with both qualitative and quantitative factors. As an example again take that of the two-way classification, where factor α is quantitative, say, and where factor β is qualitative. The combination of the two types of models (2-5) and (2-7) leads to the model of the present case:

$$y_{\alpha\beta\rho} = mu_0 + \sum_{\mu=1}^{A-1} \beta_a^{(\mu)} X_{a\alpha}^{\mu} + \sum_{\nu=1}^{B-1} b_\nu u_\nu + \sum_{\mu=1}^{A-1} \left\{ \sum_{\nu=1}^{B-1} \beta_\nu^{(\mu)} W_{\nu\alpha}^{\mu} \right\} + e_{\alpha\beta\rho} \qquad (2\text{-}8)$$

where

$$\left.\begin{aligned} W_{\nu\alpha}^{\mu} &= W_\nu^{\mu}(\alpha) = u_\nu X_a^{\mu}(\alpha) = u_\nu X_{a\alpha}^{\mu} \\ \text{and } (\beta &= \text{level number of factor } \beta!) \\ u_\nu &= \begin{cases} 1 \text{ if } \beta = \nu \\ 0 \text{ if } \beta \neq \nu \end{cases} \end{aligned}\right\} \qquad (2\text{-}9)$$

For example, with A=B=3, the W_ν^{μ} represent the interaction terms in the following manner: W_1 and W_2 are the DIVs of the interaction effect $\alpha_{linear} \times \beta$, and W_1^2 and W_2^2 are the DIVs of $\alpha_{quadratic} \times \beta$.

Summarizing, there are three types of DIVs in the analysis of variance part of the NOVACOM model, where each DIV represents an individual degree of freedom of an ANOVA effect: (1) the u_ν and their products representing individual degrees of freedom of qualitative factorial effects, (2) the X_a^{μ}, X_b^{ν}, ... and their products, representing quantitative factorial effects, and (3) the W_ν^{μ}, representing individual degrees of freedom of interaction effects between qualitative and quantitative factors. All factorial effects involving at least one qualitative factor with more than two levels are represented by groups of 2 or more DIVs. These are the groups of DIVs which are subjected to the backward ranking procedure as will be discussed in Section 2.1.2.

When the given data layout contains unoccupied cells, or "empty" cells, it is not always possible to fit the constants of the ANOVA part of the model in a unique way. In other words, in case of empty cells some of the factorial effects may be confounded. In Appendix A a method is described for fitting the constants when the model is to

contain interaction effects and when there are empty cells. The method includes rules for fitting the constants of the various possible models in case of the presence of confounded effects.

In the case of confounded effects, the user of NOVACOM has the possibility to analyze all possible models, and to search for the "most probable" significant model. Each one of the possible models is treated as a separate problem by NOVACOM since the design matrix and the matrix of the normal equations (summation matrix) must necessarily be different for each model, or set of constants fitted. The various problems corresponding to the various possible models are, therefore, referred to as "Set No. v", v = 1,2,3,.... Since the input of these different sets of constants is done via Control Card No. 4 of the program, the sets are also referred to as "Control Card 4 Set No. v", v = 1,2,3,.... For the use of the results from the various sets see Sections 3.1.3 and 3.3.2.

For the case of an analysis of covariance model, NOVACOM provides another option to change the model. The author has shown (Abt [1960], pp. 102/103) for the case of one covariate (T=1 in the present notation), in which way the analysis of covariance results are related to those of the analysis of variance (excluding the covariate) when the factors and their interactions exercise significant effects upon the covariate. (The covariate then, naturally, does not fulfill the condition of being a "fixed variate"; but one has to face this situation which often occurs in practice.) In fact, in analysis of covariance, if the factors have significant effects upon the covariate, the significance of the factorial effects, with respect to the dependent variable, may be considerably reduced when compared to the case where the covariate is not included in the model. (See Example 6 in Section 3.4.6.) A corresponding situation exists, naturally, when there is more than one covariate in the model. In other words, if significant covariates are kept in the model, the analyst cannot be sure that the true significance of the factorial effects is shown in the results of the analysis of covariance.

In order to give the user a possibility to judge the significance of the factorial effects without having the significant covariates in the model, NOVACOM will optionally run an analysis of variance for the factorial effects part of the model alone, i.e., without all covariates. Also under this option, additional analyses of variance are run for all OCIVs which were significant, i.e., the significant OCIVs take the place of the dependent variable in these analyses to study the influence of the factors and their interactions upon the covariates which turned out to be significant in the analysis of covariance model. These additional analyses of variance are identified as "ANVAs" in the program and will be referred to by this name in the remainder of the present report. As to the use of the ANVAs see Section 3.3.3.

2.1.2 The Backward Ranking Method

As stated in the Introduction, the program NOVACOM is mainly intended as a tool to screen, for significant factorial effects, analysis of variance (or covariance) models for crossed classifications with incomplete and unbalanced data. The method applied to this end in NOVACOM is the "backward ranking method" discussed in Abt [1967]. By this method the individual and/or groups of independent variables of the model (2-2) are ranked in an ascending order of importance.

Speaking, for the moment, of an analysis of variance model only, i.e., of a model (2-2) without covariates (T=0), the ranking is done as follows: At the first step, that ANOVA effect (the group of DIVs) is deleted from the model which among all effects "admissible for ranking" (to be defined later) has the smallest prediction power for y as measured by its "non-significance" (also to be defined later); at the second step, those two ANOVA effects are deleted from the model which together have the smallest prediction power for y, where one of the two effects is the one ranked least important at the first step and where the second effect is an effect "admissible for ranking" at this second step; and so forth until all ANOVA effects are ranked. This method leads to a unique ranking by importance of all ANOVA effects and enables the user of the program to define, at a prechosen significance level α, a "significant model."

When the general model of analysis of covariance ($T > 0$) is again assumed, the T covariates (CIVs) are ranked in a manner corresponding to that described before for the groups of DIVs, however, in this case one independent variable is deleted from the model at each step. The ranking of the CIVs is done first, i.e., all original N-T DIVs are kept as part of the model while the T CIVs are being ranked.

The ranking process of the CIVs is abbreviated as "COMO" = "COncomitant variables Magnitude (of prediction power for y) Ordering", and the subroutine which performs COMO in the program is identically named. Correspondingly, the ranking process of the groups of DIVs is abbreviated as "FEMO" = "Factorial Effects Magnitude (of prediction power for y) Ordering", and the subroutine performing FEMO in NOVACOM is again identically named. The names COMO and FEMO are also used, in a more general meaning, to refer to the whole analysis of covariance part and to the whole analysis of variance part, respectively, of the program.

As can be seen from the above, the ranking is performed cumulatively, that is, at each step all individual and/or groups of independent variables ranked at previous steps are included in the group of independent variables sought at the present step to have minimum

prediction power for y. (The reason for ranking "cumul tively" is explained further below.) The cumulative ranking principle is based on what may be termed the "main theorem of multiple regression." The content of the theorem, see, for example, Anderson and Bancroft [1952], p. 172, is as follows:

MAIN THEOREM. Given the linear model (2-1)

$$y = \beta_0 + \sum_{\nu=1}^{N} \beta_\nu x_\nu + e,$$

the residuals, e, are assumed to be normally independently distributed with expectation zero and variance σ^2. Under the null hypothesis $H_0\{\beta_{\nu_1} = \beta_{\nu_2} = \cdots = \beta_{\nu_{N-N'}} = 0\}$, where $\{\beta_{\nu_1}, \beta_{\nu_2}, \cdots, \beta_{\nu_{N-N'}}\}$ are the regression coefficients of a specified set of N-N' independent variables whose contribution to the "total" regression sum of squares (due to all N independent variables) is to be tested, the variance ratio

$$F_c = \frac{SS_{N-N'}}{N-N'} \Big/ \frac{ATSS - ASSR(N)}{n-N-1} \qquad (2\text{-}10)$$

is distributed as F with N-N' and n-N-1 degrees of freedom. The terms in this formula are defined as follows:

ASSR(N) = "total" regression sum of squares adjusted for the mean, with N degrees of freedom, due to all N independent variables;

$SS_{N-N'}$ = ASSR(N) - ASSR(N') = "additional" regression sum of squares, with N-N' degrees of freedom, due to the specified set of N-N' independent variables, where ASSR(N') is as defined below;

ASSR(N') = regression sum of squares adjusted for the mean, with N' degrees of freedom, due to the N' < N independent variables left in the model after deleting the N-N' independent variables whose contribution to the fit is to be tested;

ATSS = $\sum_{i=1}^{n} (y_i - \bar{y})^2$ = total sum of squares (of y) adjusted for the mean, with n-1 degrees of freedom;

n = total number of observed y-values.

In the terms of the NOVACOM model, the "specified set of N-N' independent variables" equals the sum of (1) the CIV or group of DIVs to be considered for ranking at a given step and (2) all previously ranked IVs. The prediction power for y of this set may be tested by F_c of (2-10), that is, the null hypothesis that the set of N-N' IVs do not have any prediction power for y may be tested. Obviously, the more significant F_c is, the more important is the corresponding group of independent variables for y, and vice versa. This leads to the ranking criterion "Non-Significance" as used in NOVACOM for the actual ranking:

$$\text{Non-Significance} = \int_{F_c}^{\infty} \varphi(F)dF \qquad (2\text{-}11)$$

where $\varphi(F)$ is the probability density function of F with N-N' and n-N-1 degrees of freedom. The non-significance is the tail area under the density curve of F to the right of the calculated value F_c. One can easily see from (2-11) that the non-significance equals the significance level α in the test (2-10) when F_c equals $F_{1-\alpha}$.

The importance of the non-significance as a ranking criterion lies in the fact that, in comparing various sets of independent variables, the possibly varying degrees of freedom, N-N', of the sets are taken into account. The term non-significance is derived from the fact that a set of IVs with a non-significance which is larger than that of another set of IVs can be considered as having a prediction power for y which is smaller than that of the second set.

At each step of the analysis of variance part of NOVACOM, the non-significances are computed, for each admissible effect at that step, with the incomplete beta function ratio, see, for example, Greenwood and Hartley [1962], p. 182:

$$\left.\begin{aligned} \text{Non-Significance} &= \int_{F_c}^{\infty} \varphi(F)dF = I_{x_c}(\tfrac{1}{2}f_2, \tfrac{1}{2}f_1), \\ \text{where} \quad x_c &= \left(1 + \frac{f_1}{f_2} F_c\right)^{-1} \\ f_1 &= \text{N-N'} \\ f_2 &= \text{n-N-1} \end{aligned}\right\} \qquad (2\text{-}12)$$

In the text of the present report, the computed value of the non-significan will be referred to as "I(X)." The subroutine included in NOVACOM for this computation is called ISUBX and is based on a method by DiDonato and Jarnagin [1966].

One can see that in the analysis of covariance part of NOVACOM, i.e., in COMO, the ranking criterion "non-significance" is equivalent to that of the "additonal regression sum of squares", $SS_{N-N'}$, of the Main Theorem, (2-10), because at each step the degree of freedom f_1 is constant for all CIVs to be considered for ranking. Consequently, the CIVs are ranked in the program according to $SS_{N-N'}$. However, once the least important CIV at a given step has been found according to the smallest $SS_{N-N'}$, the I(X)-value (2-12) for the corresponding group of N-N' CIVs is computed in order to provide information for the determinatio of the significant CIVs.

The program defines the significant CIVs, which are to be kept in the model during the later ranking of the factorial effects, as follows. From the established ranking order in COMO (which is achieved while keeping all original N-T DIVs in the model) and the I(X)-values the program looks for the "first significant step", i.e., the step where $I(X) \leq \alpha$ for the first time. (This value α, which is specified from one of three α-values chosen by the program user, is defined as "α-value No. KALPHA"; where KALPHA = 1,2, or 3, is also the choice of the user.) All CIVs ranked before the step where $I(X) \leq$ ALPHA (KALPHA) for the first time will be deleted permanently from the model, whereas the others (i.e., the significant CIVs) will remain part of the model throughout FEMO.

The ranking of the covariates and the factorial effects in the manner described above leads to a uniquely defined orthogonal decomposition of ASSR(N), the "total regression sum of squares", into the successive "additional regression sums of squares." This is the main advantage of the ranking method compared to other methods of applying analysis of variance to incomplete and unbalanced data layouts. If all N-T degrees of freedom available in such a layout are properly ascribed to factorial effects (see also Appendix A), the regression sum of squares, ASSR(N-T), due to all N-T DIVs with which FEMO started, has degrees of freedom equal to the number of occupied cells in the layout minus one. However, it is not always desirable (or possible, due to program limitations) to ascribe all degrees of freedom "between cells" to factorial effects. In any case (whether or not ASSR(N-T) equals the sum of squares between cells), the ranking method should and will tend to ascribe a maximum portion of the regression sum of squares to a minimum number of factorial effects. Correspondingly, a maximum portion of ASSR(N) is ascribed to a minimum number of covariates and factorial effects, or, in multiple regression, to a minimum number of independent variables.

The backward direction of the ranking (as opposed to a forward direction) is necessitated by the fact that only in this way are the unities of so-called "compounds" preserved during the ranking process, see Abt [1967]. A "compound" is defined as a set of $\tilde{N}$ ($\leq$ N) IVs when the error variance $\hat{\sigma}^2$ associated with all $\tilde{N}$ IVs is smaller, by orders of magnitude, than the error variance associated with any subset of $\tilde{N}$-1 IVs, i.e., after any single IV has been excluded from the set of the $\tilde{N}$ IVs comprising the compound.

Note. The reason for ranking the CIVs and/or groups of DIVs cumulatively on the basis of F_c in (2-10) is in order to be able to maintain a valid ranking criterion through the significant model. The alternative to the cumulative ranking procedure would be to include in the numerator of a given F_c-value (2-10) the additional regression sum of squares due to only one given effect considered for ranking. For example, at the second step of the ranking, if a given effect is represented by N'-N" DIVs (where N > N' > N") one would use, instead of (2-10),

$$F_c = \frac{SS_{N'-N''}}{N'-N''} \Big/ \frac{ATSS - ASSR(N)}{n-N-1}$$

to test the null hypothesis that the N'-N" regression coefficients are all zero. However, this F_c-value is distributed as F only if the previous null hypothesis, $H_0\{\beta_{v_1} = \beta_{v_2} = \cdots = \beta_{v_{N-N'}}\} = 0$ was accepted. That is, a ranking order based on this alternative procedure would be valid only until the first significant effect is reached. From then on, that is, for all effects except the least important one contained in the significant model, the ranking order would be invalid. Since, in the method of NOVACOM, considerable emphasis is placed on the ranking order of the significant effects also, the cumulative dropping procedure based on the F_c-value (2-10) is adopted here. It is felt that the ranking order (for non-significant effects) would be changed little - if at all - if the alternative procedure would be applied. However, the significance as given by I(X) is possibly very much dependent upon whether the cumulative or the alternative ranking method is applied. For this reason, the program gives the necessary printout to provide the analyst with the information to determine the significance of the F-test at any given step according to the alternative method. See Section 3.1.3 for a more detailed discussion.

In addition to the cumulative ranking procedure, the program NOVACOM has an option to perform also "single deletion", or "single dropping" (of CIVs and/or groups of DIVs from the model). However, in the procedure of single dropping the ranking order is taken from the results of the cumulative procedure without re-employing the I(X)-criterion or sums of squares criterion for ranking. The single dropping procedure essentially consists of a redefinition of the model at each step, i.e., of a pooling of the additional regression sum of squares due to the previously ranked CIV and/or group of DIVs with the previous error sum of squares at each step. For example, at the second step of single dropping in FEMO, the

error sum of squares (which was ATSS - ASSR(N-T) at the first step) is redefined as ATSS - ASSR(N-T) + $SS^{(1)}$, where $SS^{(1)}$ is the additional regression sum of squares due to the group of DIVs representing the factorial effect which was ranked (by the cumulative dropping procedure) as least important at the first step. This means that, for the second step, the model is redefined as containing all factorial effects of the original model except the one ranked least important at the first step. For the factorial effect which was ranked second-least important at the second step of the cumulative dropping procedure, the I(X)-value (2-12) is then computed with f_1 = degrees of freedom of effect ranked second-least important, and with f_2 = n-N+T-1 + $DF^{(1)}$, where $DF^{(1)}$ are the degrees of freedom of the effect ranked least important at the first step. At the third step of single dropping in FEMO, the above degrees of freedom f_1 and f_2 are pooled (as are the corresponding sums of squares) to form the new error degrees of freedom for the third step; and so forth. The reason for computing the one I(X)-value, as indicated above, at each step of single dropping is to provide the necessary information for the determination of a significant model based on this single dropping procedure.

The reason for having the two ranking procedures in NOVACOM to determine a significant model and the use of the two procedures are discussed in Section 3.1.3.

As mentioned previously, at each step of the cumulative ranking procedure only the "admissible" effects are considered for ranking at that step. The concept of ranking under rules of "restricted admissibility" (see Abt [1967]) is based upon the fact that some of the additional regression sums of squares, $SS_{N-N'}$ of (2-10), which correspond to certain null hypotheses, are dependent upon the type of linear restrictions chosen for the model constants; see Scheffe' [1959] and Gosslee and Lucas [1965]. Scheffe', for example, has shown that the additional regression sum of squares due to any one of the two main effects, $\mathcal{A}$ and $\mathcal{B}$, in the model (2-3) of Section 2.1.1, is dependent upon the restrictions chosen for the constants a_α and b_β (α = 1,2,...,A, and β = 1,2,...,B) as long as the constants $ab_{\alpha\beta}$ of the interaction $\mathcal{AB}$ are contained in the model consisting of the N' IVs. For models with qualitative factors only the following can be shown to be generally true. The additional regression sum of squares, $SS_{N-N'}$, due to a factorial effect, is dependent upon the restrictions chosen for the constants of the model as long as a higher order interaction effect, whose symbol contains all script letters of the given factorial effect, is retained in the remaining model of the N' IVs. The given factorial effect, whose symbol consists of script letters all contained in the symbol of the higher order interaction effect, will be called a "sub-effect" of that higher order interaction effect. For example, in a three-way crossed classification with qualitative factors $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, an additional regression sum of squares, $SS_{N-N'}$, containing the effect $\mathcal{AB}$ is restriction-dependent as long as the constants of the three-

factor interaction $\mathcal{ABC}$, with respect to which $\mathcal{AB}$ is a sub-effect, are contained in the model of the N' IVs. Corresponding restriction dependences can be shown to be generally true for cases with both qualitative and quantitative factors. The relations with respect to "sub-effects", among the factorial effects in these cases will not be explicitly stated here but are implied in the admissibility rules given further below and in Section 2.2.2.

Since, in general, the type of linear restrictions is arbitrarily chosen, it is logical to look for conditions under which the additional regression sums of squares are always independent from the linear restrictions. This independence is achieved by performing the ranking procedure under so-called "restricted admissibility" rules: In the backward ranking method, a given factorial effect is considered admissible for ranking only when all effects of which the given factorial effect is a sub-effect have been deleted from the model. For example, in the two-way crossed classification (both factors qualitative) with interaction, the symbols $\mathcal{A}$ and $\mathcal{B}$ are contained in the symbol $\mathcal{AB}$ which makes $\mathcal{A}$ and $\mathcal{B}$ sub-effects of $\mathcal{AB}$. Therefore, the main effects $\mathcal{A}$ and $\mathcal{B}$ will not be considered for ranking before the interaction $\mathcal{AB}$ has been ranked (deleted). Thus, in the two-way crossed classification, ranking under restricted admissibility rules always implies ranking the interaction effect $\mathcal{AB}$ as least important. Correspondingly, in the three-way classification, the interaction effect of second order, $\mathcal{ABC}$, when fitted, will always be ranked as the least important effect. Once that is done, the interactions of first order, $\mathcal{AB}$, $\mathcal{AC}$, and $\mathcal{BC}$, become admissible for ranking. The main effect $\mathcal{A}$, for example, would become admissible only after $\mathcal{AB}$ and $\mathcal{AC}$, in addition to $\mathcal{ABC}$, had been ranked (deleted). In other words, according to the backward ranking method under restricted admissibility rules, the least important effects are always, and by definition, the interaction effects of highest order. When these interaction effects become members of the significant model, their "sub-effects" will automatically become members of the significant model too. The latter merely reflects what the statistician is always aware of: If the interaction between two factorial effects (of any order) is significant, then, in general, each one of the two factorial effects themselves is significant at least at one level (or level combination) of the other effect(s). For example, take the case of a significant interaction $\mathcal{AB}$. This significance implies, in general, that factor $\mathcal{A}$ has a significant effect upon the response variable at least for one level of factor $\mathcal{B}$; and vice versa, that factor $\mathcal{B}$ has a significant effect upon the response variable at least for one level of factor $\mathcal{A}$. (Even in "orthogonal" analysis of variance it does not make sense to conclude that "$\mathcal{AB}$ is significant but $\mathcal{A}$ and $\mathcal{B}$ are not significant", merely judging from the F-tests in the ANOVA table. Application of t-tests at the individual levels of factors $\mathcal{A}$ and $\mathcal{B}$ will, in general, show significant effects.)

In light of the above reasoning, the forced inclusion, in the significant model, of factorial effects which are sub-effects of significant interactions does not appear to be a serious drawback in the establishing of a significant model by the backward ranking method under restricted admissibility rules.

The rules of restricted admissibility in the backward ranking method are applied, in the way described above, to all qualitative factorial effects (represented by groups of DIVs of the u-type only, see Section 2.1.1). Since no linear restrictions are applied to the model constants of quantitative factorial effects, the problem of dependence upon linear restrictions does not arise with these effects. Consequently, there is, in general, no need for restricted admissibility rules in the ranking of factorial effects when all factors in the model are quantitative. The exception is when the analyst wants to arrive at a significant model which contains all polynomial terms of the quantitative factor variables having lower order than the significant terms. NOVACOM does provide an option for the indicated type of restricted admissibility rules in the ranking of effects when all factors in the model are quantitative. For example, if the term X_a^3 is significant, X_a^2 and X_a also would become terms of the significant model under this option. The option automatically also applies to the ranking of the CIVs (if any are in the model). This type of "restricted admissibility" (which actually is the name of this option for cases with quantitative factors or CIVs in the model) applies also to all cross product terms and can generally be defined as follows: Under the option of "restricted admissibility", only those CIVs or DIVs (the latter being powers or cross product terms of quantitative factor variables) are admissible for ranking at a given step which are not "Sub-CIVs" (to be defined) or "Sub-DIVs" of other CIVs or DIVs, respectively, contained in the remaining model of the N' IVs. A CIV is called a "Sub-CIV" with respect to another CIV when the symbol of the "Sub-CIV" is contained, as a factor, in the symbol of the other CIV. An obvious corresponding definition applies to "Sub-DIVs" (the DIVs being powers or product terms of quantitative factor variables). For example, $x_1x_3^2$ is a sub-CIV of x_1x_3. X_a is a sub-DIV of $X_a^2X_b$. One advantage of ranking under the option of "restricted admissibility" is that the significant model becomes invariant with respect to variable transformations (for example, when replacing x_v by $x_v-\bar{x}_v$ for reasons concerning the accuracy of the matrix inversion). For further discussion see Reference 2 and Section 3.1.3 on the use of the ranking options in the present report. (Note. The program user may choose the option "unrestricted admissibility" when he does not desire to rank the quantitative factorial effects and CIVs under the option "restricted admissibility" just described.)

In the model of the type (2-8), which contains both qualitative and quantitative factorial effects, ranking under rules of restricted admissibility will imply a logical combination of the rules outlined above separately for each one of the two types of ANOVA effects.

For example, in the two-way crossed classification, the interaction effect of the linear component of (quantitative) factor α with (qualitative) factor β will bear the symbol $X_\alpha\beta$. Under both options of restricted and unrestricted admissibility X_α and β are sub-effects of $X_\alpha\beta$ and will not become admissible for ranking until $X_\alpha\beta$ has been ranked, i.e., deleted from the model. Also, X_α^2 would not be admissible as long as $X_\alpha^2\beta$ is contained in the remaining model (consisting of N' IVs). In the same example, all interaction effects $X_\alpha\beta$, $X_\alpha^2\beta$, ..., $X_\alpha^{A-1}\beta$, will be admissible at the first step of ranking when the option "unrestricted admissibility" is chosen, which here actually means "relaxed admissibility" since the admissibility restrictions originating from the presence of the qualitative effects still do exist. In the case discussed before when the analyst wants to keep all polynomial terms of lower order (than the order of the significant ones) in the significant model, the user chooses the option for "restricted admissibility." Under this option, in the above example, at the first step only $X_\alpha^{A-1}\beta$ would be admissible, followed by $X_\alpha^{A-2}\beta$ at the second step, and so forth.

Summarizing, there are two options for admissibility when the NOVACOM model contains quantitative factors or CIVs: "restricted admissibility" according to which all lower order polynomial terms are kept in the model all the time, and "unrestricted admissibility" according to which the lower order terms are not necessarily kept in the model. When applied to the ANOVA part of the model only, "unrestricted admissibility" is referred to as "relaxed admissibility." For more details on the admissibility options see Section 2.2.

The rules for ranking factorial effects under restricted admissibility stated so far, if adhered to, assure the independence of the additional regression sums of squares, $SS_{N-N'}$, from the linear restrictions chosen when all cells of the data layout are occupied. For the model of the type (2-5) with only qualitative factorial effects, adherence to the rules assures this independence also for data layouts with empty cells. However, for the model of the type (2-8) with both qualitative and quantitative factorial effects, the established rules are not sufficient to assure the independence in case of empty cells. For this situation, this author has not yet been able to completely define the pattern of the restriction dependence of the additional regression sums of squares.

As a safeguard in this case, a procedure is used in NOVACOM which is overconservative in its restrictions on the admissibility but assures the independence of the additional regression sums of squares from the linear restrictions imposed on the model constants. The procedure essentially consists of treating certain interaction effects as if they were interactions between qualitative factors only. For a formal definition of these "partially fitted full effects", as they are called, and of the procedure indicated, see Section 2.2.2.

A final remark in this discussion of the backward ranking method concerns the fact mentioned before that the (cumulative) ranking is not terminated when the "significant model" is reached but is continued until all factorial effects have been ranked. This continued ranking through the significant model serves two purposes: (1) The analyst will obtain a ranking of the factorial effects which are contained in the significant model, i.e., he will know the relative importance of the effects in the significant model; and (2) he will get an idea of what his significant model would have looked like had he chosen a different significance level α for the determination of the significant model. The second purpose is, to a certain degree, also served by the provision given in NOVACOM to actually choose three α values for three different significant models, where a full printout of all pertinent data is given for each of these significant models.

Ranking through the significant model sometimes leads to I(X)-values which are so small that they are far beyond the accuracy limits of the subroutine ISUBX which computes I(X). In order to be able to rank the factorial effects of the significant model in this case, a provision is made in the program to automatically redefine the error sum of squares by one of three pooling procedures. These pooling procedures (marked by one, two, or three " $^{+}$"-signs attached to the step number of the ranking) increase the error sum of squares, thereby decreasing the F-value of (2-10) and increasing the I(X)-value according to (2-12). Of the three pooling procedures the first one (the " $^{+}$"-procedure) is identical to the single dropping procedure. The " $^{++}$"- and " $^{+++}$"-procedures are not justifiable from a theoretical point of view and are just "emergency" measures to ensure a complete ranking in all cases. For more details, see Sections 2.3.2 and 2.4.

2.1.3 Accuracy Checks on Matrix Inversions

Since the method of NOVACOM is based on the general linear model (2-1), the accuracy of the results is dependent upon the accuracy of the inversion of the matrix of the normal equations of rank N+1 and of all matrices of smaller rank to be inverted at subsequent steps of the ranking. The matrices may be singular (by faulty fitting of IVs) or they may be ill-conditioned such that the inverses are fictitious or inaccurate, respectively. The procedure used in NOVACOM to check on the validity of the inverses is essentially that of the program DA-MRCA (see Reference 2). The main features of the procedure are (1) computation of the matrix $I_c = A^{-1}A$, where A is the matrix of the coefficients of the normal equations and A^{-1} is its inverse, and (2) comparison of the main diagonal elements of I_c with those of the unit matrix I. If any one of the deviations $|i_{\nu\nu}-1|$ is larger than a small input value, "TOLI2", where the $i_{\nu\nu}$ are the main diagonal elements of I_c, the inverse A^{-1} is rejected and a new inversion is tried after a specified admissible CIV or admissible group of DIVs has been deleted from the model. This procedure is continued until the first

time an acceptable inverse A^{-1} (corresponding to an acceptable, or "good" model) is found. The step in the ranking method when the "first good model" is found is called the "first good step." Once the first good step has been reached, no further accuracy checks on the matrix inversions are performed. The reason is that, at the first good step, all singularities must necessarily have been eliminated, and that the accuracy of the inversion is assumed to improve with the monotonic decrease of the rank of the matrix at subsequent steps.

The reasons for checking only the main diagonal elements of the matrix I_c is derived from the fact that the off-diagonal elements of I_c are not necessarily indicative of the accuracy of the inversion; see Section VI.1.b of Reference 2.

Two preliminary checks are exercised before the procedure described above is executed: (1) a check whether the determinant of the matrix A is non-positive and (2) a check whether an element of the main diagonal of the inverse, A^{-1}, is negative. Should any one of the two events happen, the model of the corresponding step in the ranking is rejected without performing the remaining check(s).

2.1.4 Printout and Comprehensive Analyses

The program gives a "full printout" of all pertinent data at selected steps of the ranking. The selected steps include the "first good step" and each step at which a significant model according to one of the specified α-values is reached. (As mentioned, up to three such α-values may be specified of which one is to be defined as KALPHA in case both COMO and FEMO are to be run. The same set of α-values is used for both the cumulative and the single dropping procedure and in COMO as well as in FEMO.) The full printout consists essentially of the following: The elements $c_{vv'}$ of the inverse matrix, A^{-1}; the value of the determinant of A; the estimated regression coefficients, b_v, and their standard deviations, $s\sqrt{c_{vv}}$, where s is the square root of the estimated error variance for the given step; the predicted values, $\hat{Y}$, and the prediction errors, $\hat{e}$; the prediction error frequency distribution and the results of the calculation of the χ^2-test for normality on the prediction errors. The full printout for any one of the three α-values in both cumulative and single dropping should provide the program user with all information for the model he decides to use as the "significant model."

In addition to the full printout, a complete identification of the data input is printed at the beginning of each problem, including a list of the DIVs and CIVs; the observed values of the quantitative factor variables, of the QCIVs and the dependent variable(s); all averages and various other statistics; and the summation matrix (the latter consisting of the matrix A and the cross product terms with the y's).

For every step of the ranking, the I(X)-values and their arguments ($a = \frac{1}{2}f_2$; $b = \frac{1}{2}f_1$) are printed, plus an identification of the admissible CIVs or effects. The latter enables the program user to follow the ranking process in detail and to see how the program arrived at the established ranking order of CIVs and/or factorial effects.

At the end of each problem a "Final Comprehensive Analysis", ("FCA"), is printed which gives the results of the ranking in condensed form. There is an FCA printed for both cumulative and single dropping in case the option of also performing the single dropping procedure has been chosen. Each line of the FCA corresponds to one step of the ranking procedure, and the following are some of the more important items printed: The symbol of the CIV or effect ranked at this step; a "Procedure" - symbol (PRC) consisting of an asterisk when this step corresponds to one of the three specified significance levels α; the I(X)-value; the two mean squares in the F-test plus their degrees of freedom; and the coefficient of determination. See Section 3.3.1 for a detailed discussion of the FCA. The ANVAs, when applicable (i.e., when there were significant covariates in an analysis of covariance model and when the ANVA option was chosen), are also printed in the form of FCAs at the end of a problem.

When several Control Card 4 Sets have been run, a "Final FCA" is printed, repeating the FCAs from each problem in order to facilitate a convenient search for the most probable significant model. See Section 3.4.5 for an example of how to use such a Final FCA.

2.2 Automatic Generation and Controls

The present section contains a description of the automatic generation of the model terms (CIVs, DIVs, effects), the controls over the admissibility of these terms during the ranking processes, and the generation of the design matrix in NOVACOM. The notation used is that which is also printed by the program in the identification of CIVs, DIVs and factorial effects. Together with the contents of the next section (2.3), the contents of the present section originally served as the programmer's information for coding NOVACOM.

2.2.1 Generation and Admissibility of CIVs

First, the notation will be introduced which is used for the CIVs (concomitant independent variables) in the analysis of covariance part of the NOVACOM model.

OCIVs ("Original" CIVs) i.e., CIVs of which a physically observed value exists for each observed value of the dependent variable, y, are denoted by their cardinal numbers followed by a "1" in parentheses, indicating the first power, that is, the OCIV itself: 1(1), 2(1), 3(1), etc., corresponding to x_1, x_2, x_3, etc., respectively, in the usual notation.

GCIVs ("Generated" CIVs) which are powers of OCIVs are denoted by the cardinal number of the OCIV followed, in parentheses, by the power to which the OCIV is raised; for example, 2(2), 4(6), etc., corresponding to x_2^2, x_4^6, etc., respectively. GCIVs which are products of powers of OCIVs are formed by connecting CIVs by "x's", for example, 1(1) x 2(2) corresponding to $x_1x_2^2$, and 2(1) x 3(2) x 4(6) corresponding to $x_2x_3^2x_4^6$. Accordingly, the following definitions are introduced:

> A "CIV" is identified by one or more pairs of numbers, the pairs being connected by "x's". The first number in each pair is called the "cardinal number" or "OCIV-number"; the second number is put in parentheses and is called the "power."

Quite often the user will want to fit a model of order $p > 1$ in the concomitant variables. In general, this would mean that all possible terms up to order p of the OCIVs are desired in the model. For example, with two OCIVs, 1(1) and 2(1), a complete model of order p=2 would include the following 5 CIVs: 1(1), 2(1), 1(2), 2(2), 1(1) x 2(1). (In usual notation, these are x_1, x_2, x_1^2, x_2^2, x_1x_2.) The generation of a pth order model is done automatically in NOVACOM according to the order p=P put in columns 14 + 15 of Control Card No. 1. (See Section 3.1.1.) With TP (columns 12 + 13, Control Card No. 1) being the number of OCIVs, the program will generate a total number of GCIVs equal to:

$$T - TP = \sum_{j=1}^{P} \binom{TP + j - 1}{j} - TP, \qquad (2\text{-}13)$$

and the total number of CIVs will be

$$T = \sum_{j=1}^{P} \binom{TP + j - 1}{j}. \qquad (2\text{-}14)$$

In the following Table 2.1, a scheme is given for the generation of an example model of order p=3 with TP=4 OCIVs. The first four columns, one each for each OCIV, give the powers to which the OCIVs are raised to form, after multiplication, the CIV given in the fifth column. The last column gives the number of CIVs in each order-group.

CCIV No.				CIV-Symbol	
1	2	3	4		
1	0	0	0	1(1)	$\binom{4+1-1}{1} = 4$
0	1	0	0	2(1)	
0	0	1	0	3(1)	
0	0	0	1	4(1)	
2	0	0	0	1(2)	$\binom{4+2-1}{2} = 10$
1	1	0	0	1(1) x 2(1)	
1	0	1	0	1(1) x 3(1)	
1	0	0	1	1(1) x 4(1)	
0	2	0	0	2(2)	
0	1	1	0	2(1) x 3(1)	
0	1	0	1	2(1) x 4(1)	
0	0	2	0	3(2)	
0	0	1	1	3(1) x 4(1)	
0	0	0	2	4(2)	
3	0	0	0	1(3)	$\binom{4+3-1}{3} = 20$
2	1	0	0	1(2) x 2(1)	
2	0	1	0	1(2) x 3(1)	
2	0	0	1	1(2) x 4(1)	
1	2	0	0	1(1) x 2(2)	
1	1	1	0	1(1) x 2(1) x 3(1)	
1	1	0	1	1(1) x 2(1) x 4(1)	
1	0	2	0	1(1) x 3(2)	
1	0	1	1	1(1) x 3(1) x 4(1)	
1	0	0	2	1(1) x 4(2)	
0	3	0	0	2(3)	
0	2	1	0	2(2) x 3(1)	
0	2	0	1	2(2) x 4(1)	
0	1	2	0	2(1) x 3(2)	
0	1	1	1	2(1) x 3(1) x 4(1)	
0	1	0	2	2(1) x 4(2)	
0	0	3	0	3(3)	
0	0	2	1	3(2) x 4(1)	
0	0	1	2	3(1) x 4(2)	
0	0	0	3	4(3)	

Table 2.1

In addition to the automatic generation feature for CIVs, NOVACOM provides options to delete CIVs from the automatically generated set or to add CIVs to this set. The latter mode of specifying the CIV-part of the NOVACOM model may also be used to input the entire set of CIVs by hand. The individual CIVs to be deleted or generated will be put on Control Card No. 5 (see Section 3.1.1) in the notation as described above. The various options are provided to make possible the most economic generation of the CIV-part of the model. For the use of the options see Section 3.1.2.

As mentioned before, the program user has the choice between "restricted admissibility" and "unrestricted admissibility" of the CIVs in the ranking process of the CIVs in COMO. (See column 38 of Control Card No. 1, Section 3.1.1.) "Unrestricted admissibility" simply means that all CIVs not yet ranked at a given step of COMO are admissible for ranking at that step.

The restricted admissibility option is governed by the following definitions:

> At a given step of COMO only those CIVs are admissible for ranking which are not "Sub-CIVs" of other CIVs not yet ranked. A CIV is a "Sub-CIV" of another CIV when (1) for each cardinal number in the symbol of the (sub-) CIV there is the same cardinal number present in the symbol of the other CIV, and (2) the powers of the (sub-) CIV are not larger than the corresponding ones of the other CIV.

In order to illustrate the above definitions, take the example of the model of order p=3 given above. For simplification, assume that the GCIVs of order p=3 except 1(3), 2(3), 3(3), and 4(3) have been deleted from the complete set of T=34 CIVs by means of Control Card No. 5. Then, the relations between the CIVs are as given in Table 2.2. In this example, therefore, the following 10 CIVs would be admissible at the first step of COMO, under the option of restricted admissibility: 1(1) x 2(1), 1(1) x 3(1), 1(1) x 4(1), 2(1) x 3(1), 2(1) x 4(1), 3(1) x 4(1), 1(3), 2(3), 3(3), 4(3).

2.2.2 Generation of DIVs and Admissibility of Factorial Effects

The notation for the DIVs (design independent variables) in the analysis of variance part of the NOVACOM model is based on the following definitions. A DIV is defined to be of order d, where d is the order of the factorial effect which the DIV is representing (possibly together in a group with other DIVs). For example, the DIVs of main effects have order 1; the DIVs of two-factor interactions have order 2; and in general, the DIVs of interactions between d factors have order d. The symbol of a DIV contains d pairs of numbers, where the pairs are connected by "x's". Each pair of numbers stands for a factor of the crossed

1(1)	is Sub-CIV	to	1(2) 1(1) x 2(1) 1(1) x 3(1) 1(1) x 4(1) 1(3)
2(1)	is Sub-CIV	to	1(1) x 2(1) 2(2) 2(1) x 3(1) 2(1) x 4(1) 2(3)
3(1)	is Sub-CIV	to	1(1) x 3(1) 2(1) x 3(1) 3(2) 3(1) x 4(1) 3(3)
4(1)	is Sub-CIV	to	1(1) x 4(1) 2(1) x 4(1) 3(1) x 4(1) 4(2) 4(3)
1(2)	is Sub-CIV	to	1(3)
1(1) x 2(1)	is Sub-CIV	to	NONE
1(1) x 3(1)	is Sub-CIV	to	NONE
1(1) x 4(1)	is Sub-CIV	to	NONE
2(2)	is Sub-CIV	to	2(3)
2(1) x 3(1)	is Sub-CIV	to	NONE
2(1) x 4(1)	is Sub-CIV	to	NONE
3(2)	is Sub-CIV	to	3(3)
3(1) x 4(1)	is Sub-CIV	to	NONE
4(2)	is Sub-CIV	to	4(3)
1(3)	is Sub-CIV	to	NONE
2(3)	is Sub-CIV	to	NONE
3(3)	is Sub-CIV	to	NONE
4(3)	is Sub-CIV	to	NONE

Table 2.2

classification, and the first number in the pair is the cardinal number of the factor, to be called the "factor number." For example, the factor numbers 1, 2, 3, ... correspond to the usual factor symbols $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, ..., respectively. The second number in each pair is connected with the first by an asterisk when the factor is qualitative and equals the level number of that factor. The second number of the pair is connected with the first by a point when the factor is quantitative and equals the power to which the quantitative factor variable is raised. Accordingly, one has the following definitions:

First number in "factor pair" = "factor number"

Second number in "factor pair" = "level number" when factor qualitative

Second number in "factor pair" = "power" when factor quantitative.

For example, in a three-way crossed classification (with factors $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ all at three levels, say) factor No. 1 ($\mathcal{A}$) may be qualitative, factors No. 2 and No. 3 ($\mathcal{B}$ and $\mathcal{C}$) may be quantitative. Then, typical DIVs of order 1 are 1*1 and 1*2, where these two "factor pairs" are the two DIVs representing the main effect of $\mathcal{A}$, or (qualitative) factor No. 1. Also, the factor pairs 2.1 and 2.2 are the two DIVs of order 1 corresponding to the first and second power of the quantitative factor variable of factor No. 2. (In the notation of Section 2.1.1, these two DIVs stand for X_b and X_b^2, respectively.) The DIVs representing the interaction between the first and the second factor are then 1*1 x 2.1, 1*2 x 2.1, 1*1 x 2.2, and 1*2 x 2.2. DIVs of higher order are formed in a corresponding manner.

Whereas in the generation of CIVs the program user may want to generate a model which contains all CIVs up to a given order P, the user will want, in general, to generate a model containing all DIVs up to a given order d=D, say. This order, D, in general, will be equal to the number of factors in the data layout to be analyzed. Accordingly, NOVACOM provides an option to generate all DIVs up to a specified order D (columns 4-5, Control Card No. 1; see Section 3.1.1). The total number of DIVs generated under this option depends upon the numbers of levels of the factors. These numbers are input on Control Card No. 2 (see Section 3.1.1). According to the linear restriction (2-4) introduced in Section 2.1.1, the "level number" in a "factor pair" representing a qualitative factor can only go up to one less than the total number of levels of the factor. For example, in the case of 3 factors, where the first factor has 4 levels and is qualitative, and where the second and third factors have 2 and 3 levels,

respectively, and are quantitative, the second numbers in the factor pairs will go up to 3, 1, and 2, respectively. Therefore, if the specified order of the model to be automatically generated is D=2, the model generated will include the (3+1+2) + (3·1+3·2+1·2) = 17 DIVs listed in Table 2.3.

DIV - Symbol	
1*1	3 + 1 + 2 = 6 DIVs of Order 1
1*2	
1*3	
2.1	
3.1	
3.2	
1*1 x 2.1	3·1 + 3·2 + 1·2 = 11 DIVs of Order 2
1*2 x 2.1	
1*3 x 2.1	
1*1 x 3.1	
1*1 x 3.2	
1*2 x 3.1	
1*2 x 3.2	
1*3 x 3.1	
1*3 x 3.2	
2.1 x 3.1	
2.1 x 3.2	

Table 2.3

As can easily be seen, the DIVs of order d > 1 can be generated by forming all possible products among DIVs of order d=1 with unequal "factor numbers" (first numbers in the pairs).

Corresponding to the CIV generation discussed earlier, the program provides options to delete DIVs from the automatically generated set or to add DIVs to this set. The individual DIVs to be deleted or generated will be put on Control Card No. 4 in the notation described above. For the use of the generation options see again Section 3.1.2.

The following definitions, which are essential for the FEMO part of the analysis, refer to the final set of DIVs as generated after the application of the generation option(s) described above.

An "effect" (factorial effect, that is) is defined as the group of all DIVs of equal order which (1) have equal "factor numbers",

factor by factor, and (2) have equal "powers" in each (quantitative) "factor pair", pair by pair. (That is, the only quantities which vary in the symbols of the DIVs representing an "effect" are the "level numbers.")

Groups of DIVs representing effects according to the above definition are symbolized by replacing the factor pairs of the qualitative factors by the factor numbers alone. For example, the group of DIVs representing the $\mathcal{A} \times \mathcal{B}_{1,\ldots,3}$ interaction, when (qualitative) factor $\mathcal{A}$ has, say, 4 levels, that is, the 3 DIVs, 1*1 x 2.1, 1*2 x 2.1, 1*3 x 2.1, is symbolized by 1 x 2.1.

The number of DIVs in a group representing an "effect" equals the degrees of freedom of that effect.

As discussed in Section 2.1.2, the program user may choose between two options which control the admissibility of effects for ranking at a given step of FEMO: "restricted admissibility" and "relaxed admissibility." The two options concern ANOVA models which contain at least one quantitative factor, but the admissibility rules in both options cover also the case of only qualitative factors in the model. (The two options are coupled with the options for "restricted admissibility" and "unrestricted admissibility" of CIVs, respectively, and are controlled by the same program variable, "CAD", see column 38 of Control Card No. 1.)

The following definitions apply when there are no empty cells in the original and/or marginal data classifications for which model terms (DIVs) are fitted.

Under "restricted admissibility", an effect is admissible for ranking at a given step when it is not a "sub-effect" of other effects not yet ranked (i.e., of effects still contained in the model of the N' IVs). An effect is a sub-effect of another effect, (1) when for each factor number in the symbol of the (sub-) effect there is the same factor number present in the symbol of the other effect, and (2) when the powers in the quantitative factor pairs of the (sub-) effect are not larger than the corresponding ones of the other effect.

(Note. In Section 2.1.2, sub-effects represented by one DIV were also referred to as "sub-DIVs.")

For instance, in the example of the three-way crossed classification discussed earlier, where qualitative factor $\mathcal{A}$ had 4 levels and quantitative factors $\mathcal{B}$ and $\mathcal{C}$ had 2 and 3 levels, respectively, effect $\mathcal{A} \times \mathcal{C}_{1,\ldots,2}$, or 1 x 3.1, under restricted admissibility is a sub-effect of 1 x 3.2, 1 x 2.1 x 3.1, and 1 x 2.1 x 3.2. Only when the last three effects

have been ranked (i.e., deleted from the model of the N' IVs), does 1 x 3.1 become admissible for ranking. Also in this example for restricted admissibility, 3.1 is a sub-effect of 3.2, 1 x 3.1, 1 x 3.2, 1 x 2.1 x 3.1, and 1 x 2.1 x 3.2. As another example, in the three-way crossed classification with qualitative factors $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, effects 1, 2, 3, 1 x 2, 1 x 3, and 2 x 3 are sub-effects of 1 x 2 x 3. Once 1 x 2 x 3 is ranked (i.e., deleted from the model), 1 x 2, 1 x 3, and 2 x 3 become admissible. (Note that in the present example of qualitative factors only, effects of equal order cannot be sub-effects of each other.)

> Under "relaxed admissibility" an effect is admissible for ranking at a given step when it is not a sub-effect of other effects still contained in the model of the N' IVs where a "sub-effect" is now defined as follows: An effect is a sub-effect of another effect (1) when for each factor number in the symbol of the (sub-) effect there is the same factor number present in the symbol of the other effect and (2) when the powers in the quantitative factor pairs of the (sub-) effect are not larger than the corresponding ones of the other effect, and (3) when the (sub-) effect is of lower order than the other effect and (4) when the symbol of the other effect contains at least one qualitative factor.

For instance, in the example mentioned before, where qualitative factor $\mathcal{A}$ has 4 levels and quantitative factors $\mathcal{B}$ and $\mathcal{C}$ have 2 and 3 levels, respectively, effect $\mathcal{A} \times \mathcal{C}_{linear}$, or 1 x 3.1, under relaxed admissibility, is a sub-effect of 1 x 2.1 x 3.1 and 1 x 2.1 x 3.2. When the latter two effects are deleted from the model, 1 x 3.1 becomes admissible for ranking. Also under relaxed admissibility, 3.1 is a sub-effect of 1 x 3.1, 1 x 3.2, 1 x 2.1 x 3.1, and 1 x 2.1 x 3.2.

According to the definition of relaxed admissibility, in an ANOVA model containing only quantitative factors all effects are admissible for ranking at the first step and at all subsequent steps of FEMO. That is, for data classifications with only quantitative factors, "relaxed admissibility" corresponds to "unrestricted admissibility" in the ranking of CIVs in COMO.

When there are empty cells in the original and/or marginal data classifications for which DIVs are fitted, with both quantitative and qualitative factors present, the following definitions apply.

> A "full effect" is defined as the group of all those "effects" of equal order (containing both qualitative and quantitative factors) which have equal "factor numbers", factor by factor, when a complete set of DIVs has been generated and no DIV has been deleted. The number of DIVs in a "full effect", or the degrees of freedom of the full effect, is defined as the product of the d factor level numbers, each reduced by one.

For instance, in the example with A=4, B=2, C=3 discussed before, the full effect $\alpha \times \beta \times C$ is represented by (4-1)(2-1)(3-1) = 6 DIVs, i.e., has 6 degrees of freedom.

> A "partially fitted full effect" ("PFFE") is defined as being the group of all those effects of equal order which have equal factor numbers, factor by factor, of which at least one factor must be quantitative, and where at least one DIV is missing preventing the group from being a "full effect."

For instance, in the above example, $\alpha \times \beta \times C$ is a "PFFE" when the effect 1 x 2.1 x 3.2 (represented by 3 DIVs = 3 degrees of freedom) is not fitted because of the presence of empty cells.

The following admissibility rule applies, no matter whether the program user chooses the option of "restricted admissibility" or that of "relaxed admissibility":

> In case of a data classification with empty cells where the set of factorial effects contains PFFEs, an effect is admissible for ranking at a given step when it is not a sub-effect of a PFFE which is still contained in the model of the N' IVs. An effect is a "sub-effect" of a "PFFE" when for each factor number of the (sub-) effect there is the same factor number present in the PFFE, and where the order of the (sub-) effect is smaller than that of the PFFE.

According to the above definition, individual effects within a PFFE are not sub-effects of that PFFE. In the above example, where the effect 1 x 2.1 x 3.2 was assumed to be excluded from the full effect $\alpha \times \beta \times C$, the effect 1 x 2.1 x 3.1, for instance, is not a sub-effect of the PFFE $\alpha \times \beta \times C$; however, the effect 1 x 3.2, or $\alpha \times C_{\text{quadratic}}$, is a sub-effect of that PFFE.

In order to illustrate, in a combined manner, all admissibility rules defined, the example of a three-way classification from the beginning of the present section is fully discussed, where factors α, β, and C have 3 levels each and where factor α is qualitative and factors β and C are quantitative.

The 26 DIVs listed in Table 2.4 would result from the user's specification to generate a model of order D=3. In Table 2.4, 7 of the 26 DIVs are marked (by dash lines) to indicate that they have been deleted from the set by means of Control Card No. 4, assuming that the pattern of empty cells does not allow the fitting of these DIVs. (See also Appendix A.) The reduced set of 26-7 = 19 DIVs is given in Table 2.5 which also contains the grouping of the DIVs into effects and the grouping of effects into PFFEs

as applicable. (The symbols of PFFEs contain only the factor numbers, that is, they appear as if they were the symbols of effects containing only qualitative factors. In fact, this is how the PFFEs are actually treated in the definition of a sub-effect of a PFFE. The symbols for the PFFEs are only used in the present section.)

Table 2.6 contains the relations which exist among the 15 effects with respect to the definition of a "sub-effect" in restricted and relaxed admissibility. Clearly, the relations listed in Table 2.6 govern the ranking at the very first step of FEMO and an effect becomes admissible for ranking once all those effects have been ranked (deleted from the model, that is) of which the effect was a sub-effect at the first step.

DIV - Symbol

1*1
1*2
2.1
2.2
3.1
3.2

1*1 x 2.1
1*1 x 2.2
1*2 x 2.1
1*2 x 2.2
1*1 x 3.1
1*1 x 3.2
1*2 x 3.1
1*2 x 3.2 ------
2.1 x 3.1
2.1 x 3.2
2.2 x 3.1
2.2 x 3.2

1*1 x 2.1 x 3.1
1*1 x 2.1 x 3.2 ------
1*1 x 2.2 x 3.1
1*1 x 2.2 x 3.2 ------
1*2 x 2.1 x 3.1 ------
1*2 x 2.1 x 3.2 ------
1*2 x 2.2 x 3.1 ------
1*2 x 2.2 x 3.2 ------

Table 2.4

DIV	Effect	PFFE
1*1 1*2	1	
2.1	2.1	
2.2	2.2	
3.1	3.1	
3.2	3.2	
1*1 x 2.1 1*2 x 2.1	1 x 2.1	
1*1 x 2.2 1*2 x 2.2	1 x 2.2	
1*1 x 3.1 1*2 x 3.1	1 x 3.1	1 x 3
1*1 x 3.2	1 x 3.2	
2.1 x 3.1	2.1 x 3.1	
2.1 x 3.2	2.1 x 3.2	
2.2 x 3.1	2.2 x 3.1	
2.2 x 3.2	2.2 x 3.2	
1*1 x 2.1 x 3.1	1 x 2.1 x 3.1	1 x 2 x 3
1*1 x 2.2 x 3.1	1 x 2.2 x 3.1	

Table 2.5

		Restricted Admissibility	Relaxed Admissibility
1	is a sub-effect of	1 x 2.1	1 x 2.1
		1 x 2.2	1 x 2.2
		1 x 3.1	1 x 3.1
		1 x 3.2	1 x 3.2
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
2.1	is a sub-effect of	2.2	----
		1 x 2.1	1 x 2.1
		1 x 2.2	1 x 2.2
		2.1 x 3.1	----
		2.1 x 3.2	----
		2.2 x 3.1	----
		2.2 x 3.2	----
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
2.2	is a sub-effect of	1 x 2.2	1 x 2.2
		2.2 x 3.1	----
		2.2 x 3.2	----
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
		1 x 2 x 3 (PFFE)	1 x 2 x 3
3.1	is a sub-effect of	3.2	----
		1 x 3.1	1 x 3.1
		1 x 3.2	1 x 3.2
		2.1 x 3.1	----
		2.1 x 3.2	----
		2.2 x 3.1	----
		2.2 x 3.2	----
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
3.2	is a sub-effect of	1 x 3.2	1 x 3.2
		1 x 3 (PFFE)	1 x 3
		2.1 x 3.2	----
		2.2 x 3.2	----
		1 x 2 x 3 (PFFE)	1 x 2 x 3
1 x 2.1	is a sub-effect of	1 x 2.2	----
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
1 x 2.2	is a sub-effect of	1 x 2.2 x 3.1	1 x 2.2 x 3.1
		1 x 2 x 3 (PFFE)	1 x 2 x 3

Table 2.6

(Cont'd)

Table 2.6 (Cont'd)

		Restricted Admissibility	Relaxed Admissibility
1 x 3.1	is a sub-effect of	1 x 3.2	----
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
1 x 3.2	is a sub-effect of	1 x 2 x 3 (PFFE)	1 x 2 x 3
2.1 x 3.1	is a sub-effect of	2.1 x 3.2	----
		2.2 x 3.1	----
		2.2 x 3.2	----
		1 x 2.1 x 3.1	1 x 2.1 x 3.1
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
2.1 x 3.2	is a sub-effect of	2.2 x 3.2	----
		1 x 2 x 3 (PFFE)	1 x 2 x 3
2.2 x 3.1	is a sub-effect of	2.2 x 3.2	----
		1 x 2.2 x 3.1	1 x 2.2 x 3.1
		1 x 2 x 3 (PFFE)	1 x 2 x 3
2.2 x 3.2	is a sub-effect of	1 x 2 x 3 (PFFE)	1 x 2 x 3
1 x 2.1 x 3.1	is a sub-effect of	1 x 2.2 x 3.1	----
1 x 2.2 x 3.1	is a sub-effect of	NONE	NONE

Also note in Table 2.6 that an effect is not listed as a sub-effect of a PFFE when the effect is already listed as a sub-effect of the individual effects contained in the PFFE. For example, effect 1 is a sub-effect, under both options of "restricted" and "relaxed admissibility", of effects 1 x 2.1 x 3.1 and 1 x 2.2 x 3.1 which two effects together comprise the PFFE 1 x 2 x 3 (see Table 2.5). When an effect is a sub-effect of only some of the effects comprising the PFFE, the effect is listed as a sub-effect of the entire PFFE also. For example, effect 3.2 is listed as a sub-effect of 1 x 3.2 and of the PFFE 1 x 3.

According to Table 2.6, under the option of "restricted admissibility", the effects would become admissible for ranking in FEMO as follows. At the first step of FEMO, only effect 1 x 2.2 x 3.1 would be admissible and, consequently, would be ranked as the least important effect. At the second step, once 1 x 2.2 x 3.1 has been deleted, only 1 x 2.1 x 3.1 is admissible and will be ranked as second least important

effect. At the third step, i.e., after deletion of the PFFE 1 x 2 x 3, effects 1 x 2.2, 1 x 3.2, and 2.2 x 3.2 become admissible. Admissibility at the fourth step depends upon which one of the three admissible effects of the third step will have been ranked (deleted from the model); and so forth.

Under the option of "relaxed admissibility", the ranking, at the first two steps, would be the same as before. At the third step, in addition to the three effects being admissible under "restricted admissibility", the 5 effects 1 x 2.1, 1 x 3.1, 2.1 x 3.1, 2.1 x 3.2, and 2.2 x 3.1 would become admissible; etc.

2.2.3 Generation of the Design Matrix

The design matrix is defined as the matrix, with n rows and N+1 columns, of the n coordinate values of the N independent variables, augmented by a column vector of n 1's for the constant, $x_0 \equiv 1$. Speaking in terms of the model (2-1), n is the number of observations of the dependent variable, y. The number N of IVs is the sum of the number (T) of CIVs and the number (N-T) of DIVs. This implies that the presently discussed generation of the design matrix refers to the final set of IVs which enter the analysis.

The CIV-part of the design matrix is generated as follows. For each of the n observations of y there is one observation each for the TP OCIVs. The set of observations on the TP OCIVs (for each value of y) are put on Data Card No. 3, see Section 3.1.1, and enter as such (but normally coded for reasons of the accuracy of the matrix inversions, see below) into the design matrix. The GCIV-values, being powers and/or products of OCIV-values, are computed by the program, according to the specifications given by the user, and then enter the design matrix. For example, if there are TP=3 OCIVs, the set of the three numerical OCIV observations, for one selected y value from the total of n observations, may be 15, 2, and -1.1. If there is a GCIV in the model with symbol, say, 1(2) x 2(3) x 3(1), ($x_1^2 x_2^3 x_3$ in the usual notation), the program will assign to it, as a covariate value for the one selected y observation, the numerical value $15^2 \times 2^3 \times (-1.1)^1 = 225 \times 8 \times (-1.1) = -1980$.

In case the program user chooses the option for coding the OCIV values (see columns 20-23 of Control Card No. 1, Section 3.1.1) this coding will be done by NOVACOM according to the formula

$$x'_{\nu i} = \frac{x_{\nu i} - \bar{x}_\nu}{CR_\nu}; \quad \nu = 1,\ldots,TP; \quad i = 1,\ldots,n \qquad (2\text{-}15)$$

where

$$\bar{x}_y = \frac{1}{n} \sum_{i=1}^{n} x_{y_i},$$

$$R_y = \text{MAX}(x_{y_i}) - \text{MIN}(x_{y_i}),\ i = 1,\ldots,n,$$

and where C is an arbitrary constant usually chosen as C=1. (For a discussion of this coding, see Abt et al. [1966], Section VII.2.a., and Section 3.1.3. of the present report.) For example, if n=6, and if the CCIV x_1 has the values 15, 6, 2, 18, 10, and 3, one would have

$$\bar{x}_1 = \frac{1}{6}(15 + 6 + 2 + 18 + 10 + 3) = 9$$

$$R_1 = 18 - 2 = 16$$

and the first coded value would read (if C=1):

$$x'_{11} = \frac{15-9}{1 \cdot 16} = \frac{3}{8} = 0.375.$$

The program, under the coding option, uses the coded values to generate the GCIV values.

The DIV-part of the design matrix is generated from the information contained in the cell-identification which is given as input for each one of the n y-observations, see the 1st Data Card, Section 3.1.1. The cell identification consists of the level numbers of the cell to which the y-value corresponds. For instance, in the example of Table 2.5 (Section 2.2.2) the cell defined by α=1, β=3, and γ=1 has the cell identification "131." In accordance with the derivations in Section 2.1.1, NOVACOM assigns values to the DIVs of first order (i.e., to DIVs representing main effects) and then generates all DIVs of order d > 1 by multiplication.

A DIV of order 1 for a qualitative factor is assigned, as numerical value, a 1 when the level number of the DIV equals the corresponding level number in the cell identification. If the two level numbers are unequal, the DIV is assigned the value zero. In the example of Table 2.5, y_{131}, as mentioned, corresponds to cell 131. Accordingly, for this observation of y, DIV 1*1 receives a 1, DIV 1*2 receives a zero.

A DIV of order 1 for a quantitative factor is assigned the numerical value of that level whose number is given in the cell identification, and the level value is raised to the power of the DIV. (The level values of the quantitative factor variables are input on Control Card No. 6, see

Section 3.1.1.) For instance, in the example of Table 2.5 the numerical values of the levels of (quantitative) factor No. 2 may be -0.15, -0.01, and +0.20. For the value y_{131} then, DIV 2.1 is assigned the numerical value 0.20, and DIV 2.2 is assigned the value $(0.20)^2$. If the three levels of (quantitative) factor No. 3 are -0.50, +0.10, and +0.45, say, DIVs 3.1 and 3.2 are assigned, for again observation y_{131}, the values -0.50 and $(-0.50)^2$, respectively.

As example for the assignment of numerical values to DIVs of order $d > 1$ by multiplying the values of the respective DIVs of order 1, see the following table of selected DIVs (all DIV-values again for observation y_{131}):

DIV	Assigned Value for Observation y_{131}
1*1 x 2.1	$1 \cdot 0.20 = 0.20$
1*2 x 2.1	$0 \cdot 0.20 = 0$
1*1 x 2.2	$1 \cdot (0.20)^2 = 0.04$
1*2 x 2.2	$0 \cdot (0.20)^2 = 0$
1*1 x 3.1	$1 \cdot (-0.50) = -0.50$
. . . . .	. . . . .
1*1 x 2.2 x 3.1	$1 \cdot (0.20)^2 \cdot (-0.50) = -0.02$

The coding option of NOVACOM applies also to the values of the quantitative factor variables, and the coding formulae are similar to (2-15), for example, for quantitative factor variable X_b one has:

$$X'_{b\beta} = \frac{X_{b\beta} - \bar{X}_b}{R_b}, \quad \beta = 1,\ldots,B \qquad (2\text{-}16)$$

where

$$\bar{X}_b = \frac{1}{B} \sum_{\beta=1}^{B} X_{b\beta},$$

and

$$R_b = \mathrm{MAX}(X_{b\beta}) - \mathrm{MIN}(X_{b\beta}).$$

For instance, if X_b has the B=4 levels, say, 10, 23, 37, and 82, $\bar{X}_b$ will be 38 and R_b will be 72. The first level would then have the coded value (10-38)/72 = -0.3889. Under the coding option, DIVs of order d > 1 are generated by NOVACOM using the coded values of the DIVs of order 1.

The generated design matrix plus the n observations each on the possibly up to 4 different dependent variables, y, form the "data matrix." From the data matrix the program generates the matrix of the sums of the cross-products ("summation matrix") for the N+1 IVs (including the column vector of 1's) and the dependent variable(s). For the algebraic representation of the summation matrix see Herring [1967].

The summation matrix consists of the matrix ("A") of the normal equations with rank N+1 and the column vectors whose elements are the cross-product terms with y.

2.3 The Ranking Subroutines of NOVACOM

In this section the backward ranking subroutines COMO and FEMO are described in detail for the ranking of CIVs and of groups of DIVs, respectively. For simplification, the matrix A of the normal equations for the model containing all N IVs is assumed to have been successfully inverted, that is, the first step in the ranking process is also considered to be the first "good step." The consequences of deviations from this assumption can be seen without difficulty by following the flow charts given in Section 2.4.1. (In the flow charts, COMO and FEMO are given in loop representations. This and the fact that the possibility of rejected models is included in the flow charts account for some differences in the notations used in the present section and in the flow charts.) When the matrix inversions of one or more steps in COMO or FEMO had to be rejected on the grounds of the accuracy criteria imposed, the principal methods of COMO and FEMO, as outlined below, remain unchanged.

2.3.1 COMO

First, the option for only cumulative dropping of CIVs will be described, then, the option for the additional single dropping procedure. The description of COMO ("COncomitant variables Magnitude [of prediction power for y] Ordering") is given in terms of a general step No. h, where h = 1,2,...,T. (Since T CIVs are assumed in the model, the total number of steps in COMO is identical to the total number of CIVs in the model.) At Step No. h, h-1 CIVs will have been ranked, that is, will have been dropped from the model and are no longer contained in the set of the N' IVs; see the Main Theorem in Section 2.1.2. The dropping of CIVs from the model is synonymous with the deletion of the corresponding rows and columns from the matrix A of the normal equations. It is also

assumed that the program user has decided upon one of the two options for the admissibility of CIVs: "restricted" or "unrestricted admissibility", as previously discussed.

Step No. h of COMO, cumulative dropping:

(1) Determine the admissible CIVs of this step (No. h).

(2) Invert the matrix A of the normal equations of this step with rank N+1-(h-1) = N-h+2. (If h=1, this is the matrix of the full model with rank N+1 containing the constant, N-T DIVs, and T CIVs. If h > 1, this is the matrix from which the rows and columns corresponding to the h-1 CIVs previously ranked have been deleted.)

(3) Compute, for all admissible CIVs (x_ν's) of this step, the terms

$$SS(h,\nu) = \frac{[b_\nu^{(h)}]^2}{c_{\nu\nu}^{(h)}}$$

where

$b_\nu^{(h)}$ = regression coefficient of x_ν at step h,

$c_{\nu\nu}^{(h)}$ = main diagonal element (corresponding to x_ν) of the inverse matrix, A^{-1}, with rank N-h+2.

(4) Find, for all admissible CIVs x_ν, MIN[SS(h,ν)] = SS(h,-), and denote the CIV for which SS(h,ν) is minimum as x_h. The CIV x_h is the hth least important CIV. Store its symbol for the Final Comprehensive Analysis.

(5) Compute

$$SS(1)^{(h)} = \sum_{i=1}^{h} SS(i,-)$$

$$*DF(1)^{(h)} = h$$

$$SS(2)^{(h)} = ATSS - ASSR(N) = SS(2)^{(1)} = \text{const.}$$

$$*DF(2)^{(h)} = n-N-1 = DF(2)^{(1)} = \text{const.}$$

$$*COEFF\ DET^{(h)} = \frac{ASSR(N) - \sum_{i=1}^{h-1} SS(i,-)}{ATSS},$$

and define

$$*\text{DIFF MS}^{(h)} = SS(h,-)$$

$$*\text{DIFF DF}^{(h)} = 1 = \text{const.},$$

where the notation used is that of the Main Theorem, Section 2.1.2. Store the terms marked by asterisks for the Final Comprehensive Analysis.

(6) Using the terms in (5), compute

$$*MS(1)^{(h)} = \frac{SS(1)^{(h)}}{DF(1)^{(h)}}$$

$$*MS(2)^{(h)} = \frac{SS(2)^{(h)}}{DF(2)^{(h)}} = MS(2)^{(1)} = \text{const.}$$

$$*F^{(h)} = \frac{MS(1)^{(h)}}{MS(2)^{(h)}}$$

$$*I(X)^{(h)} = I(X^{(h)}, \frac{DF(2)^{(h)}}{2}, \frac{DF(1)^{(h)}}{2}),$$

where

$$X^{(h)} = \frac{1}{1 + \frac{SS(1)^{(h)}}{SS(2)^{(h)}}}$$ is the value denoted as x_0 in (2-12).

Store the terms marked by asterisks for the Final Comprehensive Analysis.

(7) Go to Step No. (h+1) by replacing, in the above computations (1) through (6), the index h by h+1.

Final Comprehensive Analysis of COMO, cumulative dropping.

The Final Comprehensive Analysis (FCA) of COMO, cumulative dropping, contains for each step the 9 values marked by asterisks in (5) and (6) above. Also in the FCA, each step is identified by its number, and the symbol of the CIV is given which was ranked at this step. There is one more column in the FCA, marked "PRC", in which an asterisk is printed for that step of COMO when $I(X) \leq \alpha$ for the first time occurs where α may assume up to three different specified input values so that asterisks may possibly be printed at 3 different steps of COMO, cumulative dropping. For the use of the FCA see Section 3.3.1.

COMO, single dropping.

In the following, the single dropping procedure of COMO is described in the terms of the cumulative dropping procedure given before. (The general step number, h, is the same as in COMO, cumulative dropping. In the flow charts, see Section 2.4.1, in the single dropping procedure of COMO, the general step number is denoted for clarity as "p.") When the program user chooses the option for "cumulative and single dropping", (column 24, Control Card No. 1, see Section 3.1.1), the COMO single dropping procedure will be performed in addition to the cumulative procedure. In other words, the cumulative dropping procedure is always executed. COMO, single dropping, uses the ranking order of the CIVs established by the cumulative ranking procedure. See also Section 2.1.2 for a more extensive discussion.

Step No. h of COMO, single dropping:

(1) Compute

$$SS(1)^{(h)} = SS(h,-)$$

$$*DF(1)^{(h)} = 1 = \text{const.}$$

$$SS(2)^{(h)} = ATSS - ASSR(N) + \sum_{i=1}^{h-1} SS(i,-)$$

$$*DF(2)^{(h)} = n-N+h-2$$

$$*COEFF\ DET(h) = \frac{ASSR(N) - \sum_{i=1}^{h-1} SS(i,-)}{ATSS} = \frac{ATSS - SS(2)^{(h)}}{ATSS},$$

and define

$$*DIFF\ MS^{(h)} = SS(h,-)$$

$$*DIFF\ DF^{(h)} = 1$$

Store the terms marked by asterisks for the FCA.

(2) Use the terms in (1) and compute $MS(1)^{(h)}$, $MS(2)^{(h)}$, $F^{(h)}$ and $I(X)^{(h)}$ as shown in (6) of Step h, COMO, cumulative dropping, and store them for the FCA.

(3) Go to Step No. (h+1) by replacing in the above computations (1) and (2), the index h by h+1.

The Final Comprehensive Analysis of COMO, single dropping, corresponds in all features to the FCA of COMO, cumulative dropping. Therefore, no further discussion is necessary.

2.3.2 FEMO

After terminating COMO (if applicable), the program goes to the subroutine FEMO ("Factorial Effects Magnitude [of prediction power for y] Ordering") for the ranking of factorial effects represented by groups of DIVs. As was discussed in Section 2.1.2, the model of the first step of FEMO includes the significant CIVs from COMO if there were significant CIVs according to the α-value No. KALPHA, α_{KALPHA}. The step of COMO at which $I(X) \leq \alpha_{KALPHA}$ for the first time is referred to, in FEMO, as h_0. See also paragraphs (2) and (5) below.

The description of FEMO is given in terms of a general step No. k, where k = 1,2,..., number of step at which last effect is ranked. The dropping of a factorial effect from the model is synonymous with the deletion, (from the matrix A) of the rows and columns which correspond to the DIVs in the group of DIVs representing the factorial effect.

The following definitions, which correspond to those used in COMO, are used in the formulation of FEMO: The admissible effects at the kth step of FEMO are defined by the arguments (k,i), where i = 1,2,..., is the set of admissible effects at this step. The argument (k,i) is used, for example, in SS(k.i) = Additional Regression Sum of Squares, at the kth step, due to that group of DIVs which represent admissible effect "i." The computation of SS(k,i) in paragraph (3) below is as given, for example, in Hader and Grandage [1958]. The term DF(k,i) stands for "Degrees of Freedom" of the effect with argument (k,i), or: of effect (k,i), and is equal to the number of DIVs representing effect (k,i).

If FEMO was preceded by a COMO, the same option regarding the admissibility as was chosen for COMO is applicable for FEMO when quantitative factors are contained in the ANOVA model: "restricted admissibility" of factorial effects, or "unrestricted admissibility" which here means "relaxed admissibility." Otherwise (assuming there was no COMO), it is supposed that the program user has decided upon one of the two options. The option for only cumulative dropping of groups of DIVs will be described first.

Step No. k of FEMO, cumulative dropping:

(1) Determine the admissible effects (k,i).

(2) Invert the matrix A of the normal equations with rank M_k, where

$$M_k = N+2 - h_o - \sum_{j=1}^{k-1} DF(j,-) \quad \text{if there were significant CIVs } (h_o \le T)$$

or

$$M_k = N+1 - T - \sum_{j=1}^{k-1} DF(j,-) \quad \text{if there were no significant CIVs or no CIVs at all } (T=0).$$

(Note: The argument (j,-) denotes one of the factorial effects ranked by FEMO prior to Step No. k. See paragraph(6) below.)

The above implies the inversion of a matrix which results from the original (N+1)x(N+1) matrix of the normal equations by deleting (a) the (h_o-1) or T rows and columns corresponding to the (h_o-1) or T, respectively, non-significant CIVs and (b) those $\sum_{j=1}^{k-1} DF(j,-)$ rows and columns which correspond to the DIVs representing the effects ranked at the previous k-1 steps. That is, the matrix with rank M_k contains, if applicable, the $T-h_o+1$ rows and columns corresponding to the significant CIVs from COMO.

(3) Compute for all admissible effects (k,i) of this step the values

$$SS(k,i) = \left(b_{i_1}^{(k)} \; b_{i_2}^{(k)} \; \cdots \; b_{i_\nu}^{(k)} \; \cdots \; b_{i_D}^{(k)}\right) \begin{pmatrix} c_{i_1 i_1}^{(k)} & c_{i_1 i_2}^{(k)} & \cdots & c_{i_1 i_\nu}^{(k)} & \cdots & c_{i_1 i_D}^{(k)} \\ c_{i_2 i_1}^{(k)} & c_{i_2 i_2}^{(k)} & \cdots & c_{i_2 i_\nu}^{(k)} & \cdots & c_{i_2 i_D}^{(k)} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ c_{i_\nu i_1}^{(k)} & c_{i_\nu i_2}^{(k)} & \cdots & c_{i_\nu i_\nu}^{(k)} & \cdots & c_{i_\nu i_D}^{(k)} \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ c_{i_D i_1}^{(k)} & c_{i_D i_2}^{(k)} & \cdots & c_{i_D i_\nu}^{(k)} & \cdots & c_{i_D i_D}^{(k)} \end{pmatrix}^{-1} \begin{pmatrix} b_{i_1}^{(k)} \\ b_{i_2}^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ b_{i_\nu}^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ b_{i_D}^{(k)} \end{pmatrix}$$

where

i_ν = subscript indicating one of the DIVs which represent effect (k,i), with

ν = 1,2,...,D, where D = DF(k,i)

$b_{i_\nu}^{(k)}$ = regression coefficient for DIV No. i_ν at Step No. k

$c_{i_\nu i_{\nu'}}^{(k)}$ = element of inverse matrix with rank M_k for row corresponding to DIV No. i_ν and for column corresponding to DIV No. $i_{\nu'}$.

(4) Compute, for all admissible effects (k,i) of this step, the values

$$SS(1)^{(k,i)} = \sum_{j=1}^{k-1} SS(j,-) + SS(k,i),$$

$$DF(1)^{(k,i)} = \sum_{j=1}^{k-1} DF(j,-) + DF(k,i)$$

$SS(2)^{(k)} = SS(2)^{(1)} = ATSS - ASSR(N) + \sum_{i=1}^{h_o-1} SS(i,-)$ if there were significant CIVs ($h_o \le T$), where $\sum_{i=1}^{h_o-1} SS(i,-)$ is taken from COMO;

or

$SS(2)^{(k)} = SS(2)^{(1)} = ATSS - ASSR(N) + \sum_{i=1}^{T} SS(i,-)$ if there were no significant CIVs or no CIVs at all (T=0).

$DF(2)^{(k)} = DF(2)^{(1)} = n-N-2 + h_o$ for first case above ($h_o \le T$);

or

$DF(2)^{(k)} = DF(2)^{(1)} = n-N-1 + T$ for second case above.

(5) Using the terms from above, compute

$$I(X)^{(k,i)} = I(X^{(k,i)}, \frac{DF(2)^{(k)}}{2}, \frac{DF(1)^{(k,i)}}{2})$$

where

$$X^{(k,i)} = \frac{1}{1 + \frac{SS(1)^{(k,i)}}{SS(2)^{(k)}}}$$

(6) Find, for the admissible effects (k,i) of this step $MAX_i[I(X)^{(k,i)}] = I(X)^{(k,-)}$ and let the kth least important effect (k,-) be defined by this equation, if $MAX_i[I(X)^{(k,i)}] > C_o = 10^{-8}$. (The numerical value of the constant ^1C_o has been chosen as 10^{-8} in accordance with the computational accuracy of the I(X)-subroutine. Note that, if $MAX_i[I(X)^{(k,i)}] \leq C_o$, the ranking procedure must have advanced well into the significant model since any chosen significance level α will be larger than 10^{-8}.) In this case, i.e., if $MAX_i[I(X)^{(k,i)}] > C_o$, compute and/or store the following terms for the Final Comprehensive Analysis of FEMO:

(a) Symbol of effect (k,-)

(b) $I(X)^{(k)} = I(X)^{(k,-)}$

(c) $DIFF\ MS^{(k)} = SS(k,-)/DF(k,-)$

(d) $DIFF\ DF^{(k)} = DF(k,-)$

(e) $DF(1)^{(k)} = DF(1)^{(k,-)}$

(f) $DF(2)^{(k)}$

(g) $MS(1)^{(k)} = \dfrac{SS(1)^{(k,-)}}{DF(1)^{(k)}}$

(h) $MS(2)^{(k)} = \dfrac{SS(2)^{(k)}}{DF(2)^{(k)}}$

(i) $F^{(k)} = \dfrac{MS(1)^{(k)}}{MS(2)^{(k)}}$

(j) $COEFF\ DET^{(k)} = \dfrac{ASSR(M_k-1)}{ATSS} = \dfrac{1}{ATSS}\left[ASSR(N) - \sum_{i=1}^{h_o-1} SS(i,-) - \sum_{j=1}^{k-1} SS(j,-)\right]$ if there were significant CIVs ($h_o \leq T$)

or:

$= \dfrac{ASSR(M_k-1)}{ATSS} = \dfrac{1}{ATSS}\left[ASSR(N) - \sum_{i=1}^{T} SS(i,-) - \sum_{j=1}^{k-1} SS(j,-)\right]$ if there were no significant CIVs or no CIVs at all (T=0).

If $MAX_i[I(X)^{(k,i)}] \leq C_o$, go to the " $^+$-procedure" as outlined below.

(7) Go to Step No. (k+1) by replacing, in the above computations (1) through (6), the index k by k+1.

The " $^{+}$-procedure" (Step k^{+}) of FEMO, cumulative dropping:

This modification of FEMO (the " $^{+}$-procedure") will apply only when $\underset{i}{MAX}[I(X)^{(k,i)}] \leq C_o$, where the superscript of I(X) may also read $(k^{+} + \delta, i)$, $(k^{++} + \varepsilon, i)$, etc., see further below. In the $^{+}$-procedure, the terms which were computed at Step No. k as described above are used. Therefore, the $^{+}$-procedure is also referred to as "Step k^{+}." The $^{+}$-procedure serves to increase the I(X)-values in order that the remaining factorial effects in the significant model may be ranked with respect to their relative importance. This is achieved by pooling all previously ranked effects into the experimental error, that is, by a redefinition of the model, as follows:

(1) Compute and/or define

$$SS(1)^{(k^{+},i)} = SS(k,i)$$

$$DF(1)^{(k^{+},i)} = DF(k,i)$$

$$SS(2)^{(k^{+})} = SS(2)^{(1)} + \sum_{j=1}^{k-1} SS(j,-)$$

$$DF(2)^{(k^{+})} = DF(2)^{(1)} + \sum_{j=1}^{k-1} DF(j,-)$$

Using the above four terms, compute the values $I(X)^{(k^{+},i)}$ as in paragraph (5) of Step No. k.

(2) Find, for the admissible effects (k,i) of this step (which are the same as in Step k):

$$\underset{i}{MAX}[I(X)^{(k^{+},i)}] = I(X)^{(k^{+},-)}.$$

If this maximum is greater than $C_o = 10^{-8}$, let the kth least important effect (k,-) be defined by this equation. In this case, compute and/or store, for the FCA, terms (a) - (j) as given in paragraph (6) of Step k, replacing the index k by k^{+}. In the FCA, print the symbol " $^{+}$" in the PRC column for this step. Then go to Step $(k^{+} + \delta)$, starting with $\delta=1$, as outlined below. In case of $\underset{i}{MAX}[I(X)^{(k^{+},i)}] \leq C_o$, go to the " $^{++}$-procedure" as outlined further below.

Step No. $(k^+ + \delta)$ of FEMO, cumulative dropping:

With $\delta = 1,2,\ldots$, this is a general step after the $^+$-procedure had to be applied. The experimental error, which was redefined at Step k^+, remains again constant, and the sums of squares due to the effects ranked are pooled again, as seen in paragraph (2) below.

(1) Determine the admissible effects $(k + \delta, i)$.

(2) After carrying out the computations similar to those of paragraphs (2) and (3) of Step No. k, compute and/or define, for the admissible effects $(k + \delta, i)$ of this step:

$$SS(1)^{(k^+ + \delta, i)} = \sum_{j=k}^{k+\delta-1} SS(j,-) + SS(k + \delta, i)$$

$$DF(1)^{(k^+ + \delta, i)} = \sum_{j=k}^{k+\delta-1} DF(j,-) + DF(k + \delta, i)$$

$$SS(2)^{(k^+ + \delta)} = SS(2)^{(k^+)}$$

$$DF(2)^{(k^+ + \delta)} = DF(2)^{(k^+)}$$

where the latter two right-hand terms are from paragraph (1) of Step k^+. Using the above four terms, compute the values $I(x)^{(k^+ + \delta, i)}$ as in paragraph (5) of Step No. k.

(3) Find, for the admissible effects $(k + \delta, i)$ of this step,

$$\underset{i}{MAX}[I(X)^{(k^+ + \delta, i)}] = I(X)^{(k^+ + \delta, -)}.$$

If this maximum is greater than C_o, let the $(k + \delta)^{th}$ least important effect $(k + \delta, -)$ be defined by the above equation. In this case, compute and/or store, for the FCA, terms (a) - (j) as given in paragraph (6) of Step No. k, replacing the index k by $(k^+ + \delta)$. Then go to Step No. $(k^+ + \delta + 1)$ as outlined above, i.e., by replacing the index δ by $(\delta+1)$. If $MAX[I(X)] \leq C_o = 10^{-8}$, go to Step $(k^+ + \delta)^+$, i.e., follow the procedure as outlined in Step k^+, replacing the index k by $(k^+ + \delta)$.

The " $^{++}$-procedure" (Step k^{++}) of FEMO, cumulative dropping:

This modification of FEMO (the " $^{++}$-procedure") will apply only when $MAX[I(X)^{(k^+, i)}] \leq C_o = 10^{-8}$, where the superscript of I(X) may also read $[(k^+ + \delta)^+, i]$, $[(k^{++} + \epsilon)^+, i]$, etc., see further below. The $^{++}$-procedure is also referred to as "Step k^{++}." The aim of the $^{++}$-procedure is to further increase the I(X)-values (which, at Step k^+, still were all below $C_o = 10^{-8}$) so that the remaining factorial effects in the significant model may be ranked with respect to their relative importance. For this purpose, at Step k^{++}, the sum of squares due to one of the admissible

effects is added to the error sum of squares according to the definition in paragraph (1) below. Later, in the computation of I(X), an "F-value" according to (2-10), see Section 2.1.2, is implied which contains this very same sum of squares in both the numerator and the denominator. So, actually, it is not an F-value, but the computational procedure of I(X) is employed nevertheless in order to have, also in this case and if applicable, a ranking criterion by which to establish the relative importance of the (highly) significant factorial effects.

The $^{++}$-procedure is as follows.

(1) Find, among the admissible effects of Step No. k, and using the terms SS(k,i) from Step No. k,

$$\min_i [MS^{(k,i)}] = \min_i \left[\frac{SS(2)^{(k^+)} + SS(k,i)}{DF(2)^{(k^+)} + DF(k,i)}\right] = MS^{(k,0)}.$$

The above equation defines the effect, which minimizes $MS^{(k,i)}$, by the argument (k,0).

(2) Compute and/or define

$$SS(1)^{(k^{++},i)} = SS(k,i)$$

$$DF(1)^{(k^{++},i)} = DF(k,i)$$

$$SS(2)^{(k^{++})} = SS(2)^{(k^+)} + SS(k,0)$$

$$DF(2)^{(k^{++})} = DF(2)^{(k^+)} + DF(k,0)$$

Using the above four terms, compute the values $I(X)^{(k^{++},i)}$ as in paragraph (5) of Step No. k.

(3) Find, for the admissible effects (k,i) of this step (which are still the same as in Step k):

$$\max_i [I(X)^{(k^{++},i)}] = I(X)^{(k^{++},-)}.$$

If this maximum is greater than C_0, let the kth least important effect (k,-) be defined by this equation. (Note that effect (k,-) will not necessarily be equal to (k,0).) In this case, compute and/or store for the FCA, terms (a) - (j) as given in paragraph (6) of Step k, replacing

the index k by k^{++}. In the FCA, print the symbol " $^{++}$" in the PRC column for this step. Then proceed to Step $(k^{++} + \epsilon)$, starting with ϵ=1, as outlined below. If $\underset{i}{MAX}[I(X)^{(k^{++},i)}] \leq C_0$, proceed to the " $^{+++}$-procedure" as outlined further below.

Step No. $(k^{++} + \epsilon)$ of FEMO, cumulative dropping:

With $\epsilon = 1,2,\ldots$, this is a general step after the $^{++}$-procedure had to be applied. The error sum of squares is now defined to consist of the sum of squares of the original error sum of squares pooled with the sums of squares due to all effects ranked before and at Step No. k. For $\epsilon = 1,2,\ldots$, the error sum of squares remains constant again. From Step No. k^{++}+1 on, the sums of squares due to the effects ranked are pooled again, as seen in paragraph (2) below.

(1) Determine the admissible effects $(k + \epsilon, i)$.

(2) After carrying out the computations which are similar to those of paragraphs (2) and (3) of Step No. k, compute and/or define, for the admissible effects $(k + \epsilon, i)$ of this step:

$$SS(1)^{(k^{++}+\epsilon,i)} = \sum_{j=k+1}^{k+\epsilon-1} SS(j,-) + SS(k+\epsilon, i)$$

$$DF(1)^{(k^{++}+\epsilon,i)} = \sum_{j=k+1}^{k+\epsilon-1} DF(j,-) + DF(k+\epsilon, i)$$

$$SS(2)^{(k^{++}+\epsilon)} = SS(2)^{(k)} + \sum_{j=1}^{k} SS(j,-)$$

$$DF(2)^{(k^{++}+\epsilon)} = DF(2)^{(k)} + \sum_{j=1}^{k} DF(j,-)$$

Using the above four terms, compute the values $I(X)^{(k^{++}+\epsilon,i)}$ as in paragraph (5) of Step No. k.

(3) Find, for the admissible effects $(k + \epsilon, i)$ of this step,

$$\underset{i}{MAX}[I(X)^{(k^{++}+\epsilon,i)}] = I(X)^{(k^{++}+\epsilon,-)}.$$

If this maximum is greater than C_o, let the $(k + \epsilon)^{th}$ least important effect $(k + \epsilon, -)$ be defined by the above equation. In this case, compute and/or store, for the FCA, terms (a) - (j) as given in paragraph (4) of Step No. k, replacing the index k by $k^{++} + \epsilon$. Then go to Step $(k^{++} + \epsilon + 1)$ as outlined above, i.e., by replacing the index ϵ by $(\epsilon+1)$. If $MAX[I(X)^{(k^{+++}\epsilon,1)}] \leq C_o$, go to Step $(k^{++} + \epsilon)^{+}$, i.e., proceed as in Step k^{+}, replacing the index k by $(k^{++} + \epsilon)$.

The " $^{+++}$-procedure" (Step k^{+++}) of FEMO, cumulative dropping:

This last modification of FEMO will apply only when $MAX[I(X)^{(k^{++},1)}] \leq C_o$ where the superscript of I(X) may also read $[(k^{+} + \delta)^{++}, i]$, $[(k^{++} + \epsilon)^{++}, i]$, etc. The $^{+++}$-procedure is also referred to as "Step k^{+++}." The aim of the procedure is to define, for Step No. k of the ranking, a least important effect after even the $^{++}$-procedure failed to increase the I(X)-values sufficiently.

In this case merely define effect (k,0) from Step k^{++} to be effect (k,-). Then compute and/or store for the FCA the terms (a) - (j) as given in paragraph (6) of Step k, replacing the index k by k^{+++}. In the FCA, print the symbol " $^{+++}$" in the PRC column for this step. Then proceed to Step No. $(k^{++} + \epsilon)$, starting with ϵ=1, as was described before.

FEMO, single dropping.

Finally for FEMO, the single dropping procedure will be discussed.

"Single dropping" is executed in addition to FEMO, cumulative dropping, when the appropriate option is chosen (column 24 of Control Card No. 1, see Section 3.1.1). As was correspondingly the case for COMO, single dropping, the single dropping procedure of FEMO uses the ranking order of the factorial effects established by the cumulative dropping procedure.

The single dropping procedure of FEMO then simply consists of the " $^{+}$-procedure" described before, which is followed all the way through, from the first to the last step. At appropriate places, the terms computed in the cumulative dropping procedure are used for the computations and/or for the Final Comprehensive Analysis, FEMO, single dropping. Since the ranking order has already been established in the cumulative procedure, the single dropping option never needs go into the $^{++}$- or $^{+++}$-procedures.

Final Comprehensive Analyses of FEMO.

The Final Comprehensive Analyses for both the cumulative and the single dropping procedure in FEMO correspond to those described for COMO. As was mentioned at the appropriate places, the symbols " $^{+}$", " $^{++}$", and

" +++" are printed in the PRC column when the corresponding procedure led to the ranking of the effect for which the symbol is printed. Also in the PRC column, an asterisk is printed whenever I(X) is smaller, for the first time, than one of the possibly up to three α-significance levels used as input, thus marking the "significant model" which corresponds to the respective α-level.

When there are two or more sets of Control Card No. 4, i.e., when the data of a given classification has been analyzed by fitting two or more models respectively, all FCAs of FEMO are repeated at the end of the problem in order to facilitate the search for the "most probable significant model." (See also Section 3.3.2.)

2.4 Computational Flow

2.4.1 Flowcharts

In this section the computational flow of NOVACOM is given in the form of logical flowcharts where these flowcharts reflect only the method of the program and are not expressed in the terms of a programming language. Some features which were discussed in previous sections, like the determination of the admissible CIVs or effects, are not contained in the charts. Wherever it is considered necessary for the understanding of the flowcharts, comments are provided which are listed in Section 2.4.2. The flowchart boxes for which comments are given in Section 2.4.2 have been marked by decimal classification numbers of which the first is the number of the chart and the second is the box number within the chart.

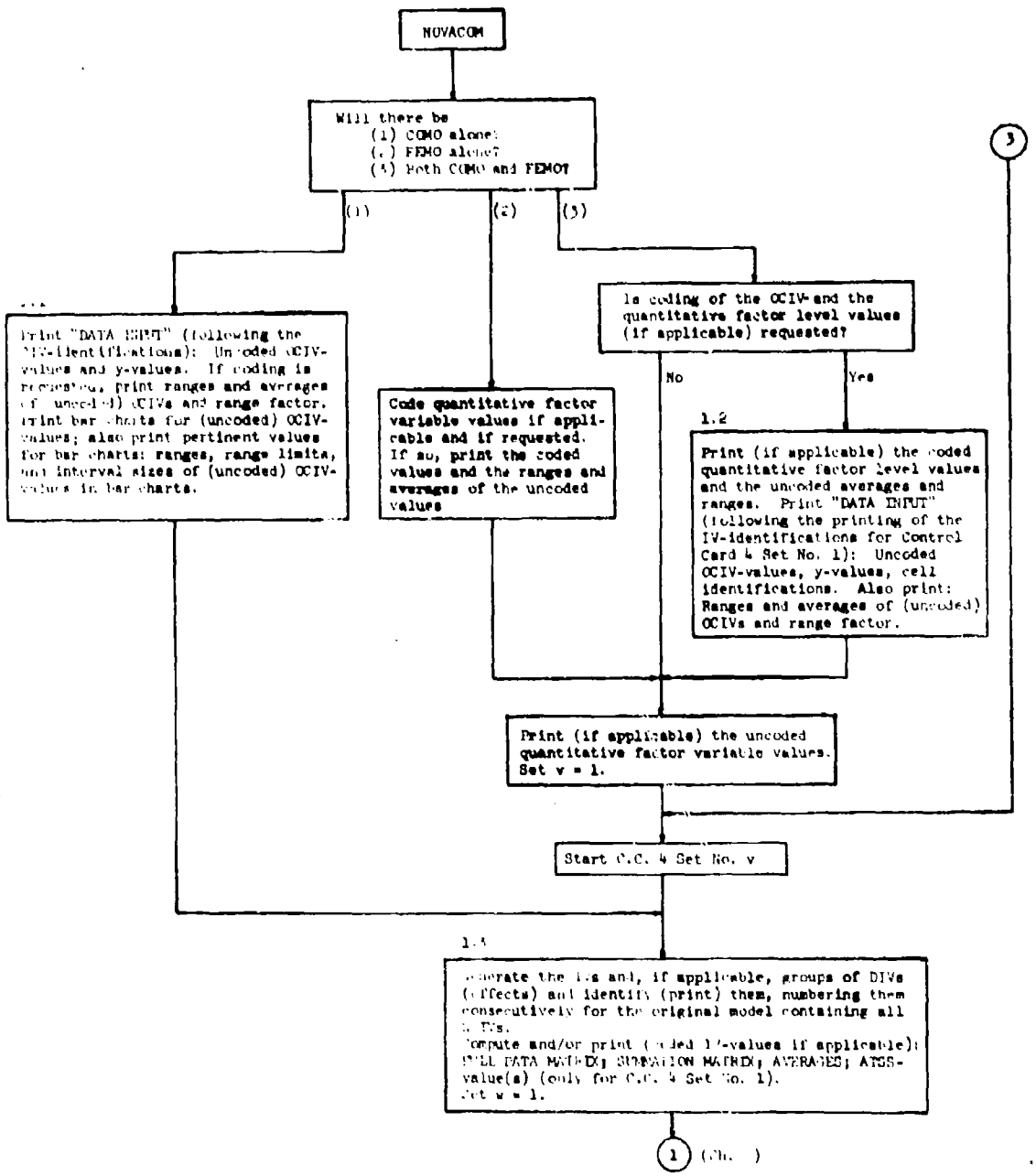

NOVACOM
Will there be
(1) COMO alone?
(2) FEMO alone?
(3) Both COMO and FEMO?
(1)
(2)
(3)
Print "DATA INPUT" (following the IV-identifications): Uncoded OCIV-values and y-values. If coding is requested, print ranges and averages of uncoded) OCIVs and range factor. Print bar charts for (uncoded) OCIV-values; also print pertinent values for bar charts: ranges, range limits, and interval sizes of (uncoded) OCIV-values in bar charts.
Code quantitative factor variable values if applicable and if requested. If so, print the coded values and the ranges and averages of the uncoded values
Is coding of the OCIV-and the quantitative factor level values (if applicable) requested?
No
Yes
1.2
Print (if applicable) the coded quantitative factor level values and the uncoded averages and ranges. Print "DATA INPUT" (following the printing of the IV-identifications for Control Card 4 Set No. 1): Uncoded OCIV-values, y-values, cell identifications. Also print: Ranges and averages of (uncoded) OCIVs and range factor.
Print (if applicable) the uncoded quantitative factor variable values. Set v = 1.
Start C.C. 4 Set No. v
1.3
Generate the IVs and, if applicable, groups of DIVs (effects) and identify (print) them, numbering them consecutively for the original model containing all IVs.
Compute and/or print (coded IV-values if applicable); FULL DATA MATRIX; SUBMATRIX; AVERAGES; ATSS-value(s) (only for C.C. 4 Set No. 1).
Set w = 1.
1 (Ch.)
3

Chart 1

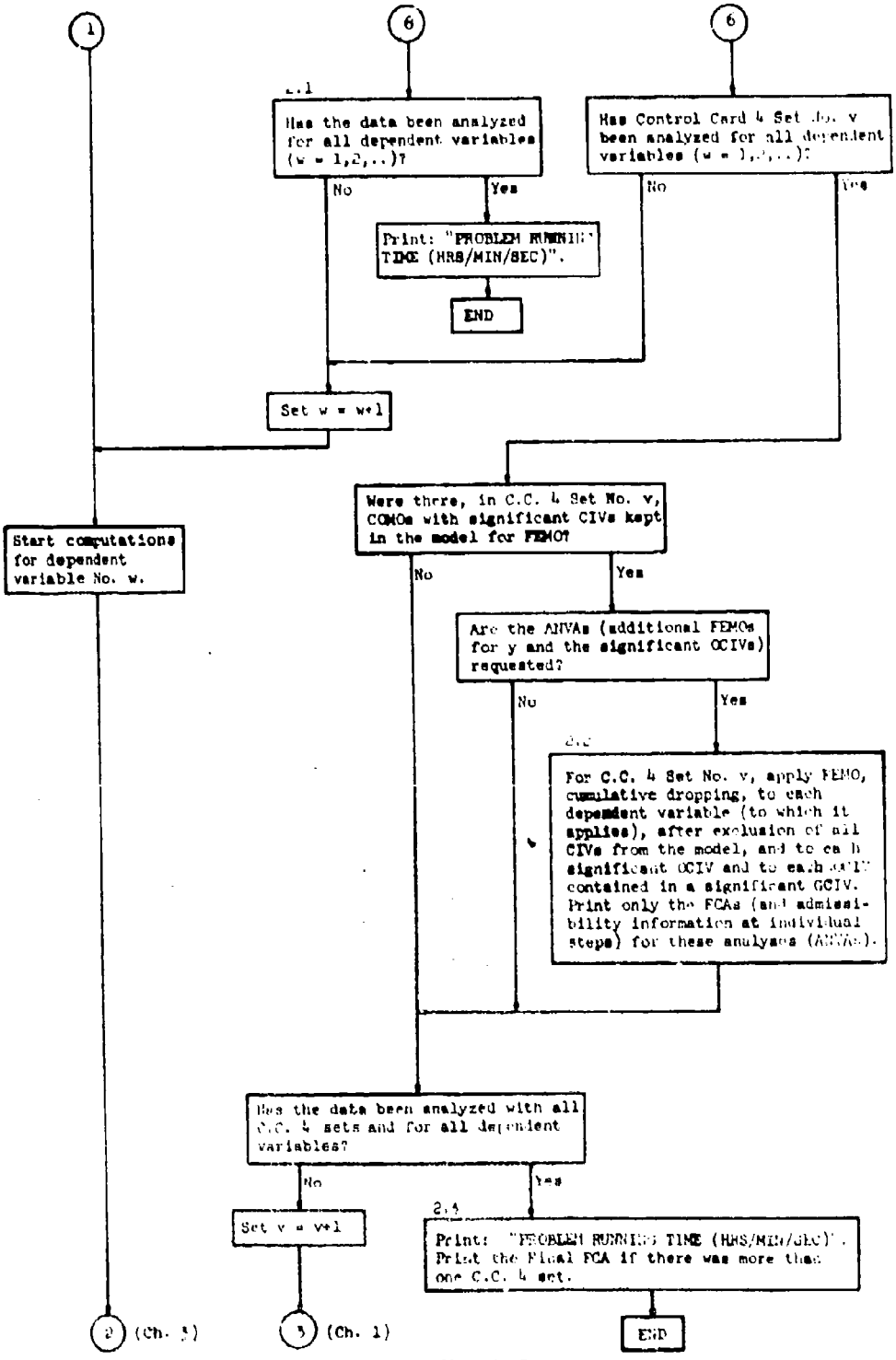

1
6
6
2.1
Has the data been analyzed for all dependent variables (w = 1,2,..)?
No
Yes
Print: "PROBLEM RUNNING TIME (HRS/MIN/SEC)".
END
Has Control Card 4 Set No. v been analyzed for all dependent variables (w = 1,2,..)?
No
Yes
Set w = w+1
Were there, in C.C. 4 Set No. v, COMOs with significant CIVs kept in the model for FEMO?
No
Yes
Start computations for dependent variable No. w.
Are the AIVAs (additional FEMOs for y and the significant OCIVs) requested?
No
Yes
2.2
For C.C. 4 Set No. v, apply FEMO, cumulative dropping, to each dependent variable (to which it applies), after exclusion of all CIVs from the model, and to each significant OCIV and to each GCIV contained in a significant GCIV. Print only the FCAs (and admissibility information at individual steps) for these analyses (AIVAs).
Has the data been analyzed with all C.C. 4 sets and for all dependent variables?
No
Yes
Set v = v+1
2.4
Print: "PROBLEM RUNNING TIME (HRS/MIN/SEC)". Print the Final FCA if there was more than one C.C. 4 set.
END
2 (Ch. 3)
3 (Ch. 1)

Chart 2

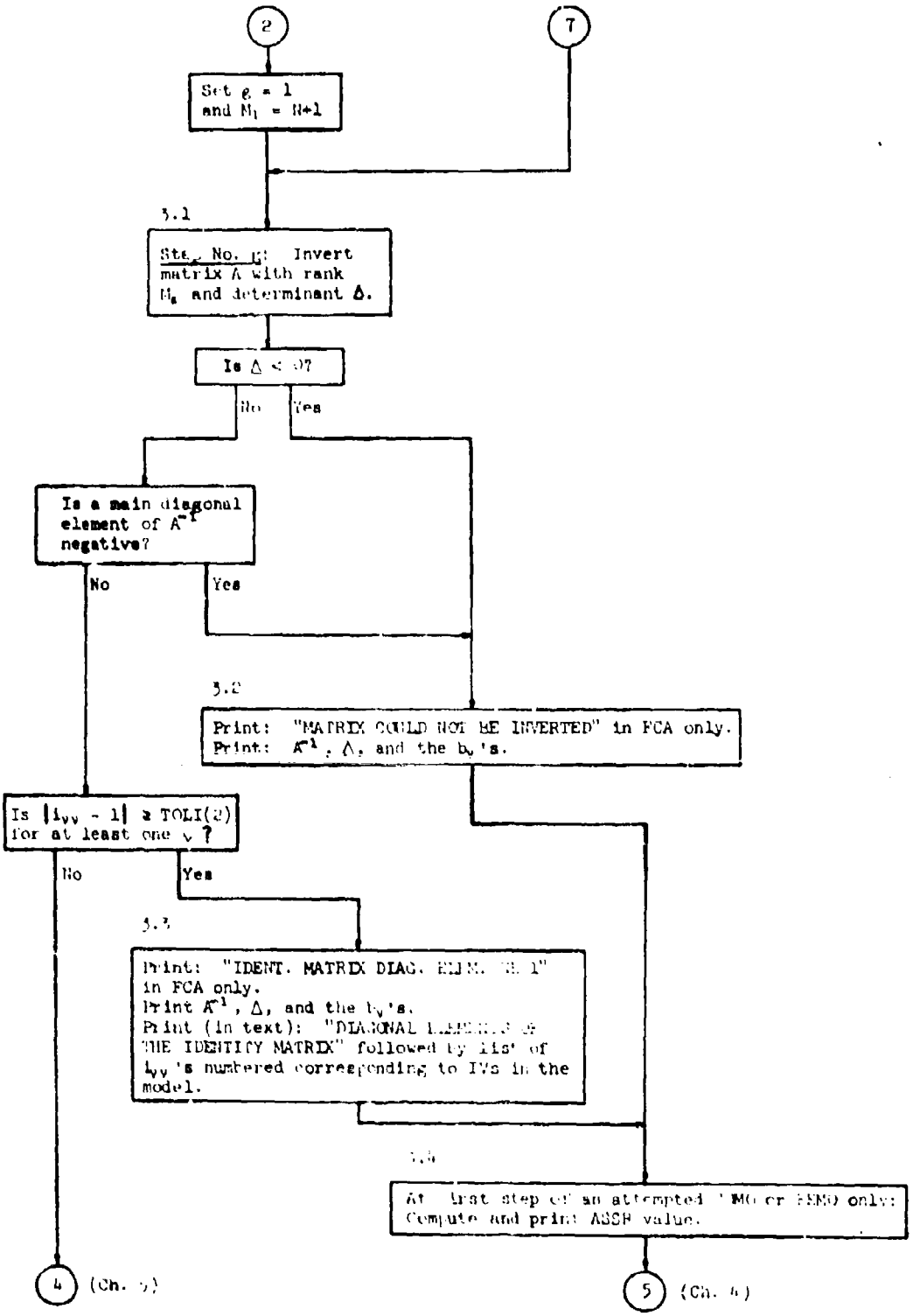

2
7
Set g = 1
and M_1 = N+1
3.1
Step No. g: Invert
matrix A with rank
M_g and determinant Δ.
Is Δ < 0?
No
Yes
Is a main diagonal
element of A^{-1}
negative?
No
Yes
3.2
Print: "MATRIX COULD NOT BE INVERTED" in FCA only.
Print: A^{-1}, Δ, and the b_ν's.
Is $|i_{\nu\nu} - 1| \geq$ TOLI(2)
for at least one ν ?
No
Yes
3.3
Print: "IDENT. MATRIX DIAG. ELEM. NE 1"
in FCA only.
Print A^{-1}, Δ, and the b_ν's.
Print (in text): "DIAGONAL ELEMENTS OF
THE IDENTITY MATRIX" followed by list of
$i_{\nu\nu}$'s numbered corresponding to IVs in the
model.
At first step of an attempted FMO or FEMO only:
Compute and print ASSR value.
4
(Ch. 5)
5
(Ch. 4)

Chart 3

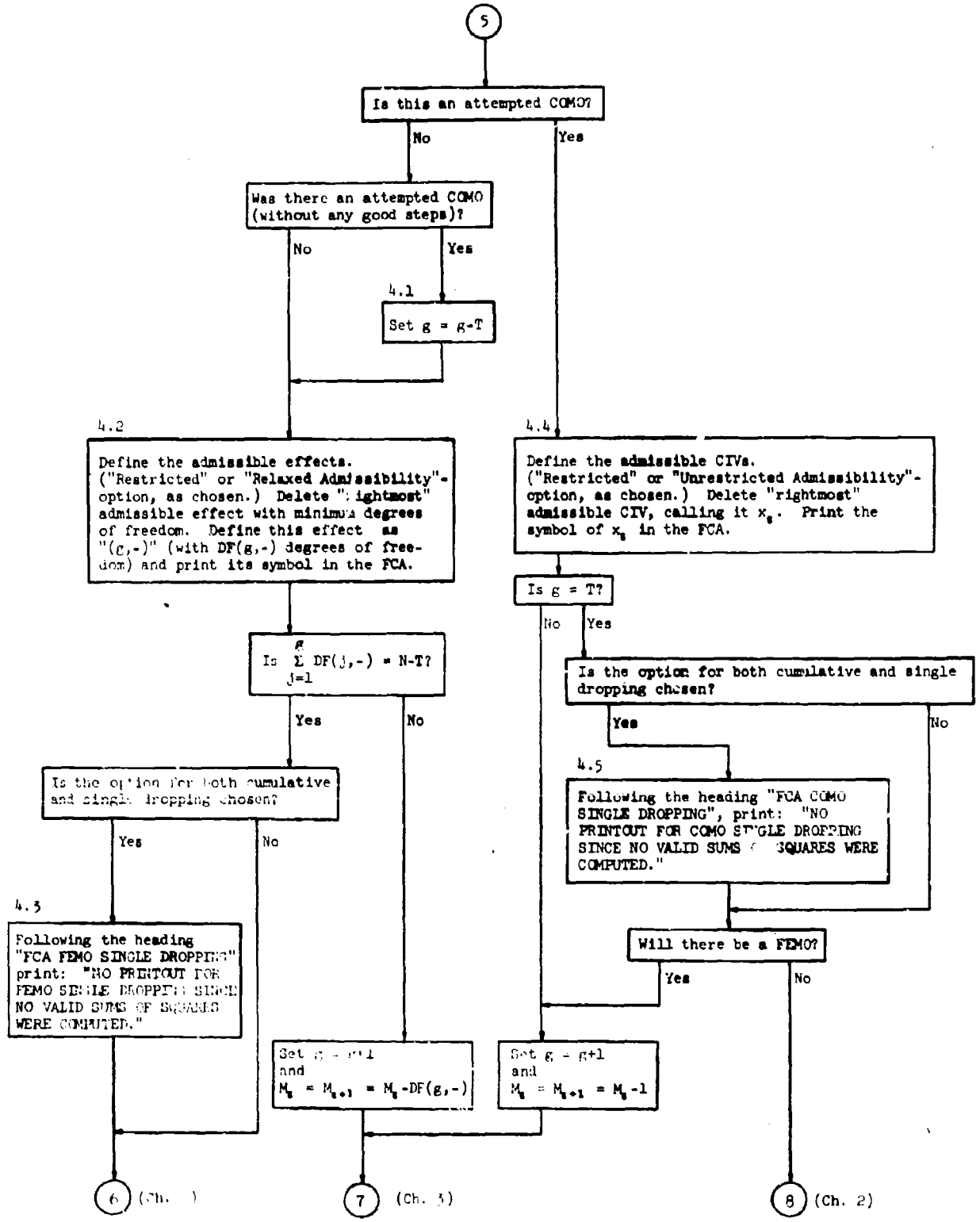

5
Is this an attempted COMO?
No
Yes
Was there an attempted COMO (without any good steps)?
No
Yes
4.1
Set g = g-T
4.2
Define the admissible effects. ("Restricted" or "Relaxed Admissibility"-option, as chosen.) Delete "rightmost" admissible effect with minimum degrees of freedom. Define this effect as "(g,-)" (with DF(g,-) degrees of freedom) and print its symbol in the FCA.
4.4
Define the admissible CIVs. ("Restricted" or "Unrestricted Admissibility"-option, as chosen.) Delete "rightmost" admissible CIV, calling it x_g. Print the symbol of x_g in the FCA.
Is $\sum_{j=1}^{g} DF(j,-) = N-T$?
Yes
No
Is g = T?
No
Yes
Is the option for both cumulative and single dropping chosen?
Yes
No
Is the option for both cumulative and single dropping chosen?
Yes
No
4.5
Following the heading "FCA COMO SINGLE DROPPING", print: "NO PRINTOUT FOR COMO SINGLE DROPPING SINCE NO VALID SUMS OF SQUARES WERE COMPUTED."
4.3
Following the heading "FCA FEMO SINGLE DROPPING" print: "NO PRINTOUT FOR FEMO SINGLE DROPPING SINCE NO VALID SUMS OF SQUARES WERE COMPUTED."
Will there be a FEMO?
Yes
No
Set g = g+1 and $M_g = M_{g+1} = M_g - DF(g,-)$
Set g = g+1 and $M_g = M_{g+1} = M_g - 1$
6 (Ch.)
7 (Ch. 3)
8 (Ch. 2)

Chart 4

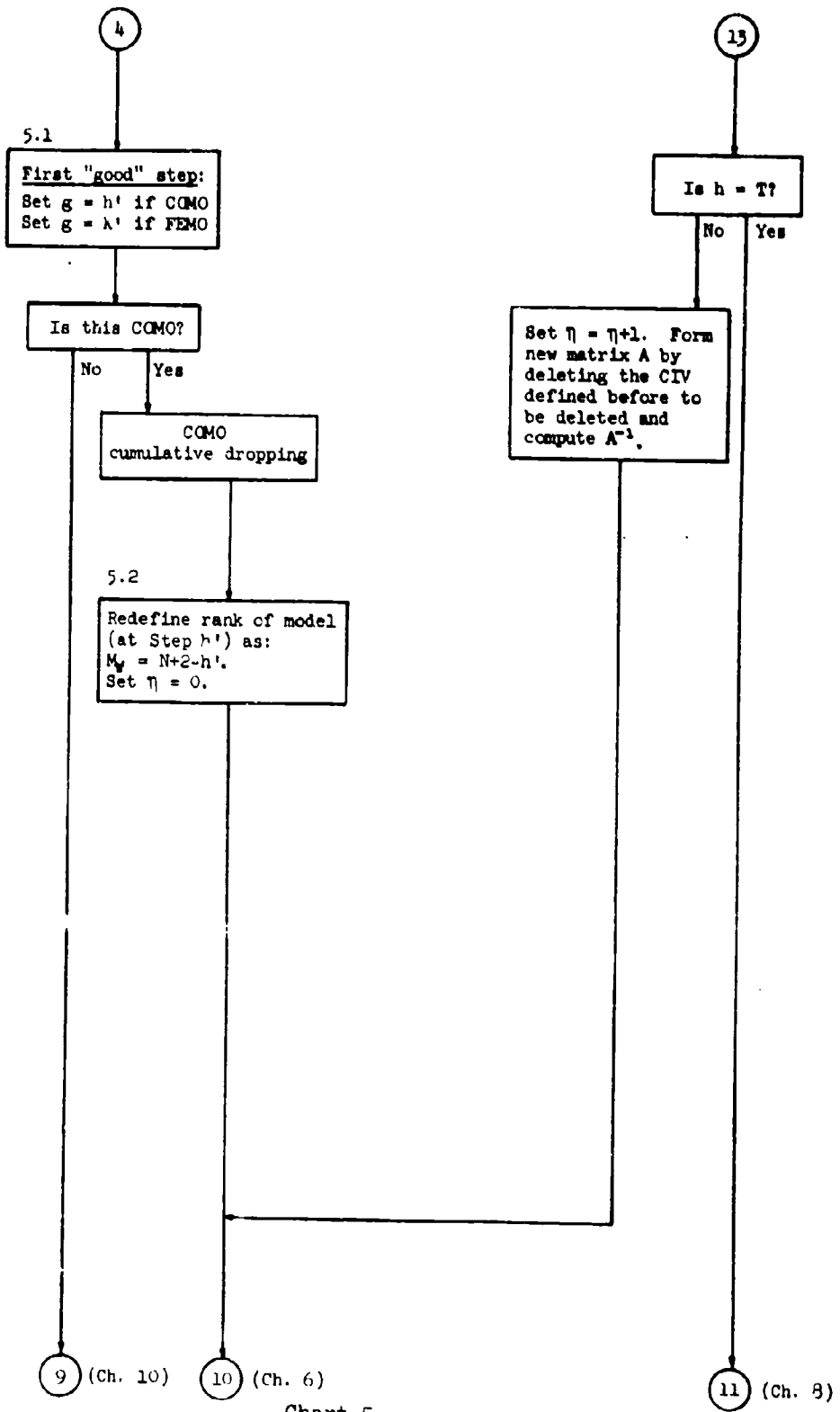

4
5.1
First "good" step:
Set g = h' if COMO
Set g = λ' if FEMO
Is this COMO?
No
Yes
COMO
cumulative dropping
5.2
Redefine rank of model
(at Step h') as:
M_{η} = N+2-h'.
Set η = 0.
13
Is h = T?
No
Yes
Set η = η+1. Form
new matrix A by
deleting the CIV
defined before to
be deleted and
compute A^{-1}.
9 (Ch. 10)
10 (Ch. 6)
11 (Ch. 8)

Chart 5

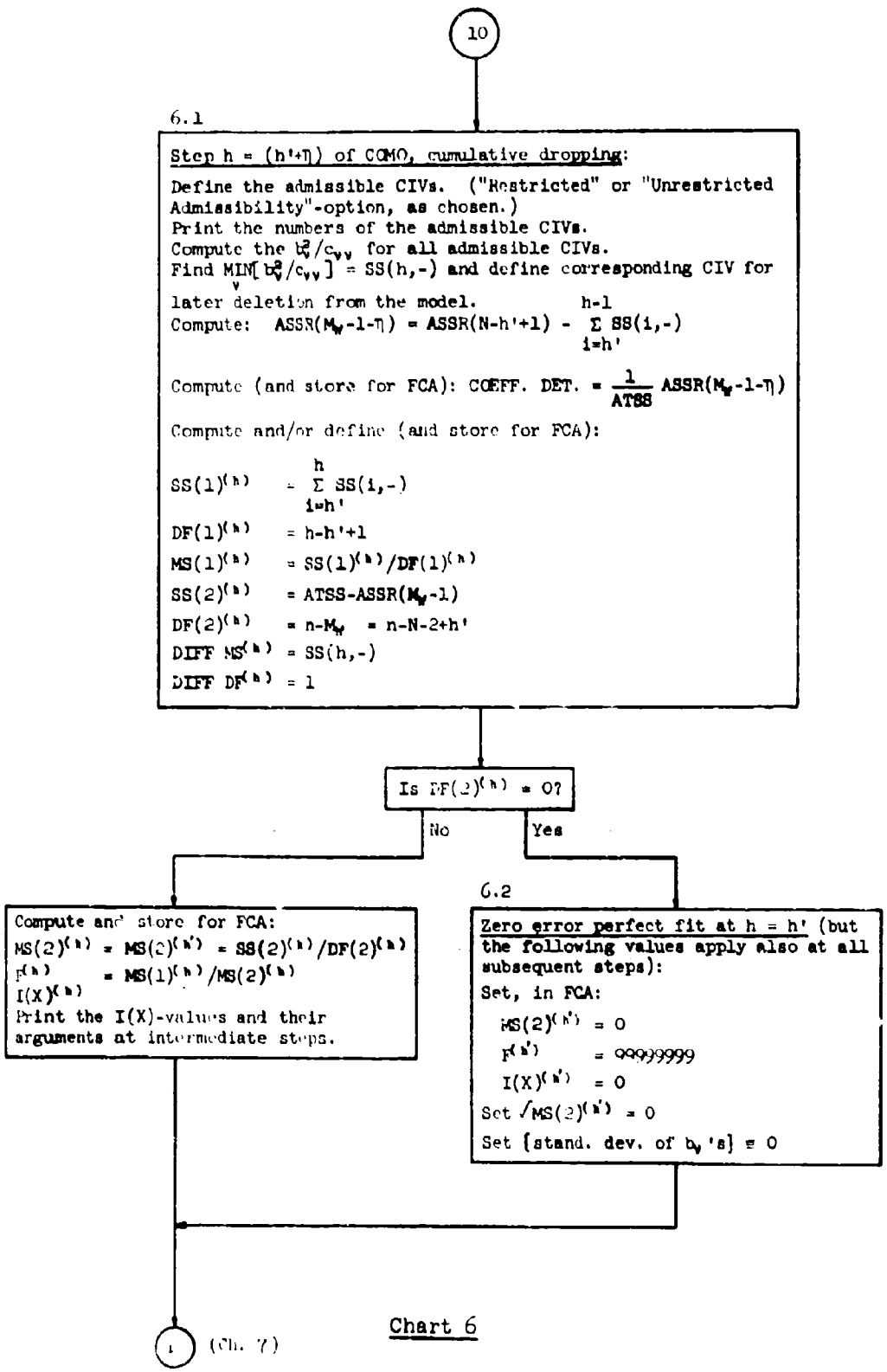

10
6.1
Step h = (h'+η) of COMO, cumulative dropping:
Define the admissible CIVs. ("Restricted" or "Unrestricted Admissibility"-option, as chosen.)
Print the numbers of the admissible CIVs.
Compute the b_v^2/c_{vv} for all admissible CIVs.
Find $MIN_v[b_v^2/c_{vv}] = SS(h,-)$ and define corresponding CIV for later deletion from the model.
Compute: $ASSR(M_w-1-\eta) = ASSR(N-h'+1) - \sum_{i=h'}^{h-1} SS(i,-)$
Compute (and store for FCA): COEFF. DET. $= \frac{1}{ATSS} ASSR(M_w-1-\eta)$
Compute and/or define (and store for FCA):
$SS(1)^{(h)} = \sum_{i=h'}^{h} SS(i,-)$
$DF(1)^{(h)} = h-h'+1$
$MS(1)^{(h)} = SS(1)^{(h)}/DF(1)^{(h)}$
$SS(2)^{(h)} = ATSS-ASSR(M_w-1)$
$DF(2)^{(h)} = n-M_w = n-N-2+h'$
$DIFF\ MS^{(h)} = SS(h,-)$
$DIFF\ DF^{(h)} = 1$
Is $DF(2)^{(h)} = 0$?
No
Yes
Compute and store for FCA:
$MS(2)^{(h)} = MS(2)^{(h')} = SS(2)^{(h)}/DF(2)^{(h)}$
$F^{(h)} = MS(1)^{(h)}/MS(2)^{(h)}$
$I(X)^{(h)}$
Print the I(X)-values and their arguments at intermediate steps.
6.2
Zero error perfect fit at h = h' (but the following values apply also at all subsequent steps):
Set, in FCA:
$MS(2)^{(h')} = 0$
$F^{(h')} = 99999999$
$I(X)^{(h')} = 0$
Set $\sqrt{MS(2)^{(h')}} = 0$
Set {stand. dev. of b_v's} = 0
1
(Ch. 7)

Chart 6

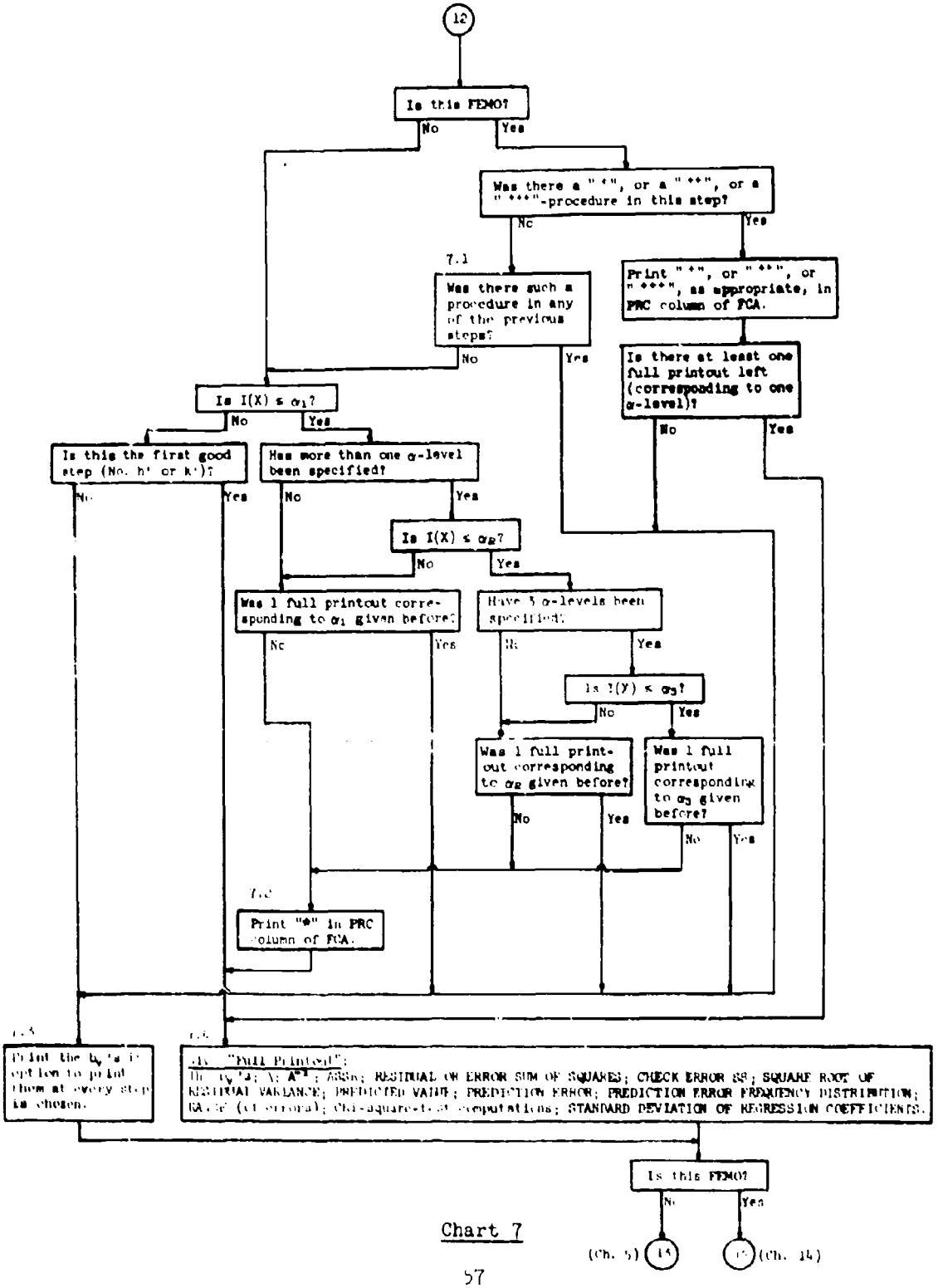

12
Is this FEMO?
No
Yes
Was there a "+", or a "++", or a "+++"-procedure in this step?
No
Yes
7.1
Was there such a procedure in any of the previous steps?
Print "+", or "++", or "+++", as appropriate, in PRC column of FCA.
No
Yes
Is there at least one full printout left (corresponding to one α-level)?
No
Yes
Is I(X) ≤ α1?
No
Yes
Is this the first good step (No. h' or k')?
Has more than one α-level been specified?
No
Yes
No
Yes
Is I(X) ≤ α2?
No
Yes
Was 1 full printout corresponding to α1 given before?
Have 3 α-levels been specified?
No
Yes
No
Yes
Is I(X) ≤ α3?
No
Yes
Was 1 full printout corresponding to α2 given before?
Was 1 full printout corresponding to α3 given before?
No
Yes
No
Yes
7.2
Print "*" in PRC column of FCA.
Print the bv's if option to print them at every step is chosen.
"Full Printout":
RESIDUAL OR ERROR SUM OF SQUARES; CHECK ERROR SS; SQUARE ROOT OF RESIDUAL VARIANCE; PREDICTED VALUE; PREDICTION ERROR; PREDICTION ERROR FREQUENCY DISTRIBUTION; (of errors); Chi-square-test computations; STANDARD DEVIATION OF REGRESSION COEFFICIENTS.
Is this FEMO?
No
Yes
(Ch. 5)
(Ch. 14)

Chart 7

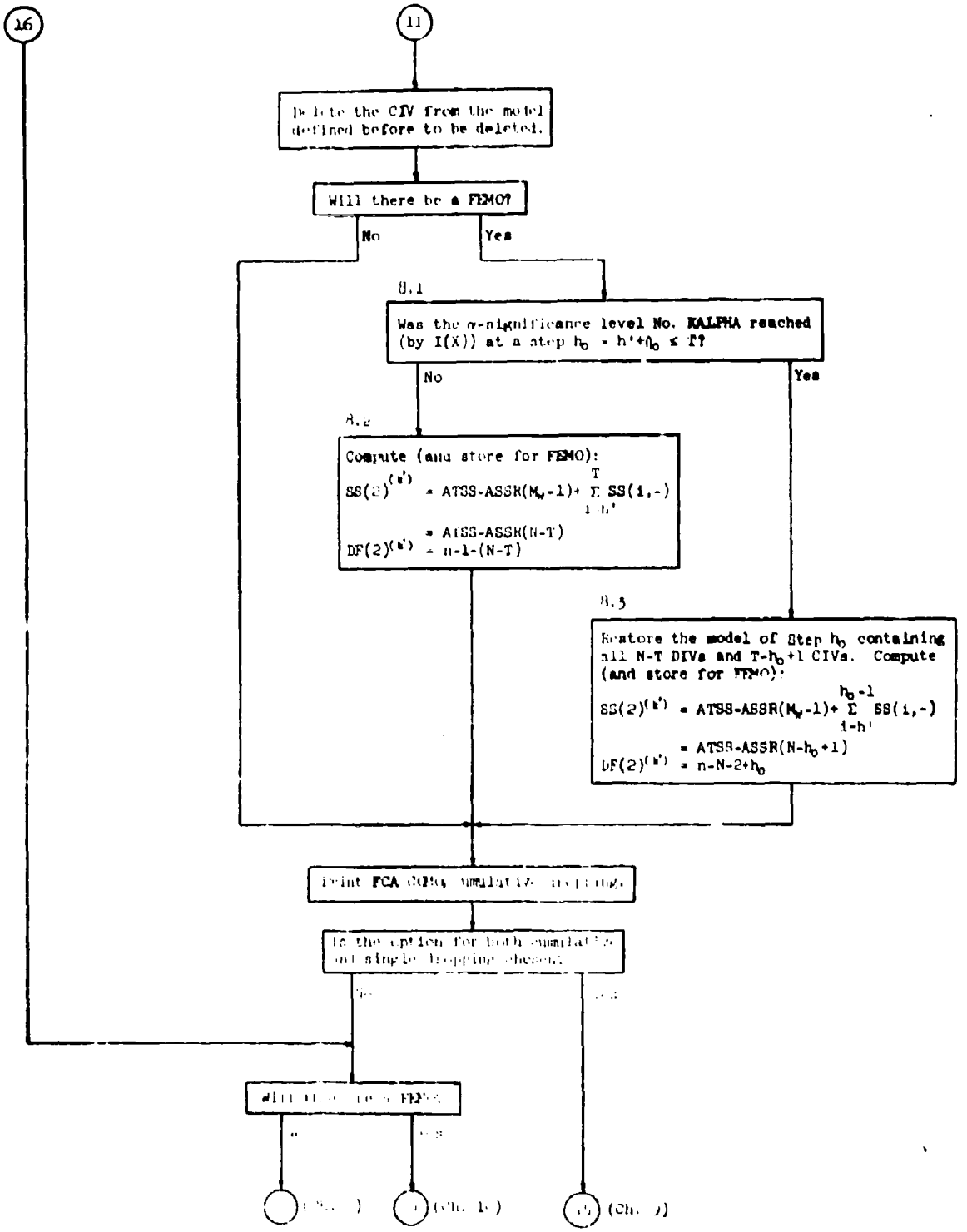

16
11
Delete the CIV from the model defined before to be deleted.
Will there be a FEMO?
No
Yes
8.1
Was the α-significance level No. KALPHA reached (by I(X)) at a step $h_0 = h'+\ell_0 \leq T$?
No
Yes
8.2
Compute (and store for FEMO):
$SS(2)^{(h')} = ATSS-ASSR(M_v-1)+\sum_{i=h'}^{T} SS(i,-)$
$= ATSS-ASSR(N-T)$
$DF(2)^{(h')} = n-1-(N-T)$
8.3
Restore the model of Step h_0 containing all N-T DIVs and $T-h_0+1$ CIVs. Compute (and store for FEMO):
$SS(2)^{(h')} = ATSS-ASSR(M_v-1)+\sum_{i=h'}^{h_0-1} SS(i,-)$
$= ATSS-ASSR(N-h_0+1)$
$DF(2)^{(h')} = n-N-2+h_0$
Print FCA edits, cumulative stepping.
Is the option for both cumulative and simple stepping chosen?
Will there be a FEMO?
(Ch. 6)
(Ch. 7)

Chart 3

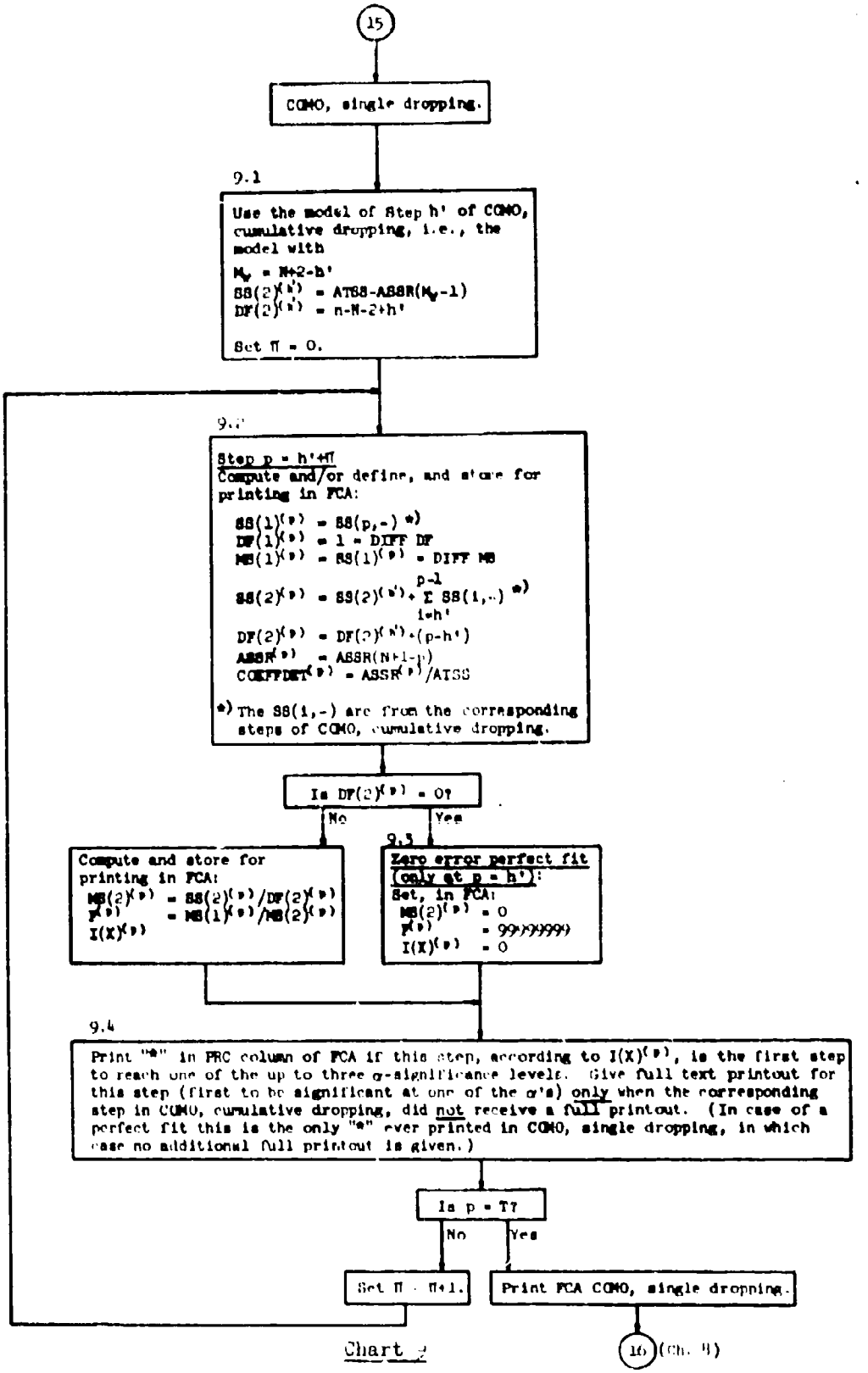

15
COMO, single dropping.
9.1
Use the model of Step h' of COMO, cumulative dropping, i.e., the model with
$M_y = M+2-h'$
$SS(2)^{(h')} = ATSS-ASSR(M_y-1)$
$DF(2)^{(h')} = n-M-2+h'$
Set Π = 0.
9.2
Step p = h'+Π
Compute and/or define, and store for printing in FCA:
$SS(1)^{(p)} = SS(p,-)$ *)
$DF(1)^{(p)} = 1$ = DIFF DF
$MS(1)^{(p)} = SS(1)^{(p)}$ = DIFF MS
$SS(2)^{(p)} = SS(2)^{(h')} + \sum_{i=h'}^{p-1} SS(i,-)$ *)
$DF(2)^{(p)} = DF(2)^{(h')} + (p-h')$
$ASSR^{(p)} = ASSR(N+1-p)$
$COEFFDET^{(p)} = ASSR^{(p)}/ATSS$
*) The SS(i,-) are from the corresponding steps of COMO, cumulative dropping.
Is $DF(2)^{(p)} = 0$?
No
Yes
Compute and store for printing in FCA:
$MS(2)^{(p)} = SS(2)^{(p)}/DF(2)^{(p)}$
$F^{(p)} = MS(1)^{(p)}/MS(2)^{(p)}$
$I(X)^{(p)}$
9.3
Zero error perfect fit (only at p = h'):
Set, in FCA:
$MS(2)^{(p)} = 0$
$F^{(p)}$ = 99999999
$I(X)^{(p)} = 0$
9.4
Print "*" in PRC column of FCA if this step, according to $I(X)^{(p)}$, is the first step to reach one of the up to three α-significance levels. Give full text printout for this step (first to be significant at one of the α's) only when the corresponding step in COMO, cumulative dropping, did not receive a full printout. (In case of a perfect fit this is the only "*" ever printed in COMO, single dropping, in which case no additional full printout is given.)
Is p = T?
No
Yes
Set Π = Π+1.
Print FCA COMO, single dropping.
16 (Ch. 8)

Chart 9

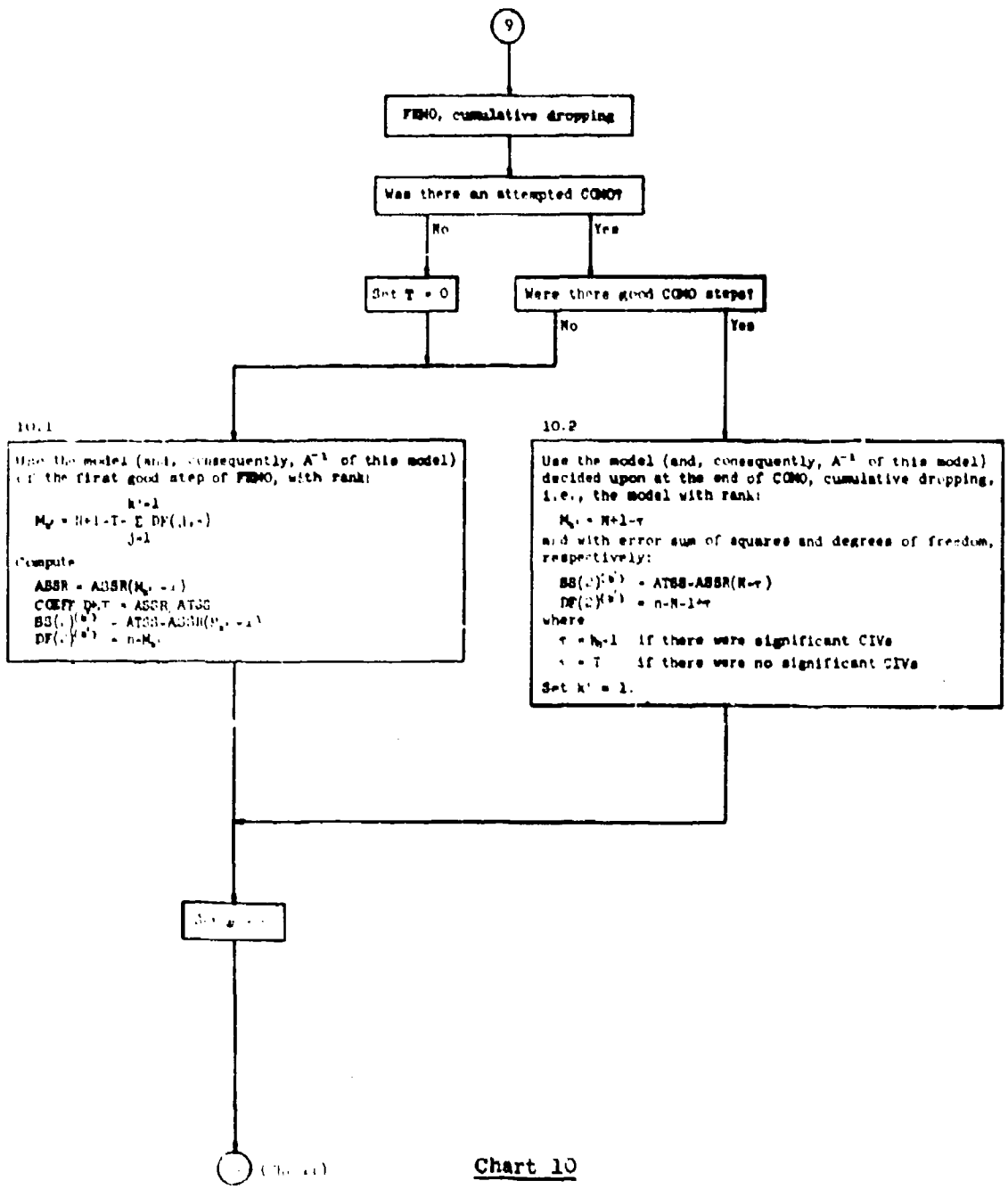

9
FEMO, cumulative dropping
Was there an attempted COMO?
No
Yes
Set T = 0
Were there good COMO steps?
No
Yes
10.1
Use the model (and, consequently, A^{-1} of this model) of the first good step of FEMO, with rank:
$M_{k'} = N+1-T-\sum_{j=1}^{k'-1} DF(j,\cdot)$
Compute
ABSR = ABSR($M_{k'}$)
COEFF DET = ASSR, ATSS
$SS(\cdot)^{(k')}$ = ATSS-ABSR($M_{k'}$)
$DF(\cdot)^{(k')}$ = n-$M_{k'}$
10.2
Use the model (and, consequently, A^{-1} of this model) decided upon at the end of COMO, cumulative dropping, i.e., the model with rank:
$M_{k'} = N+1-\tau$
and with error sum of squares and degrees of freedom, respectively:
$SS(\cdot)^{(k')}$ = ATSS-ABSR($N-\tau$)
$DF(\cdot)^{(k')}$ = n-N-1+τ
where
$\tau = h_0-1$ if there were significant CIVs
$\tau = T$ if there were no significant CIVs
Set k' = 1.

Chart 10

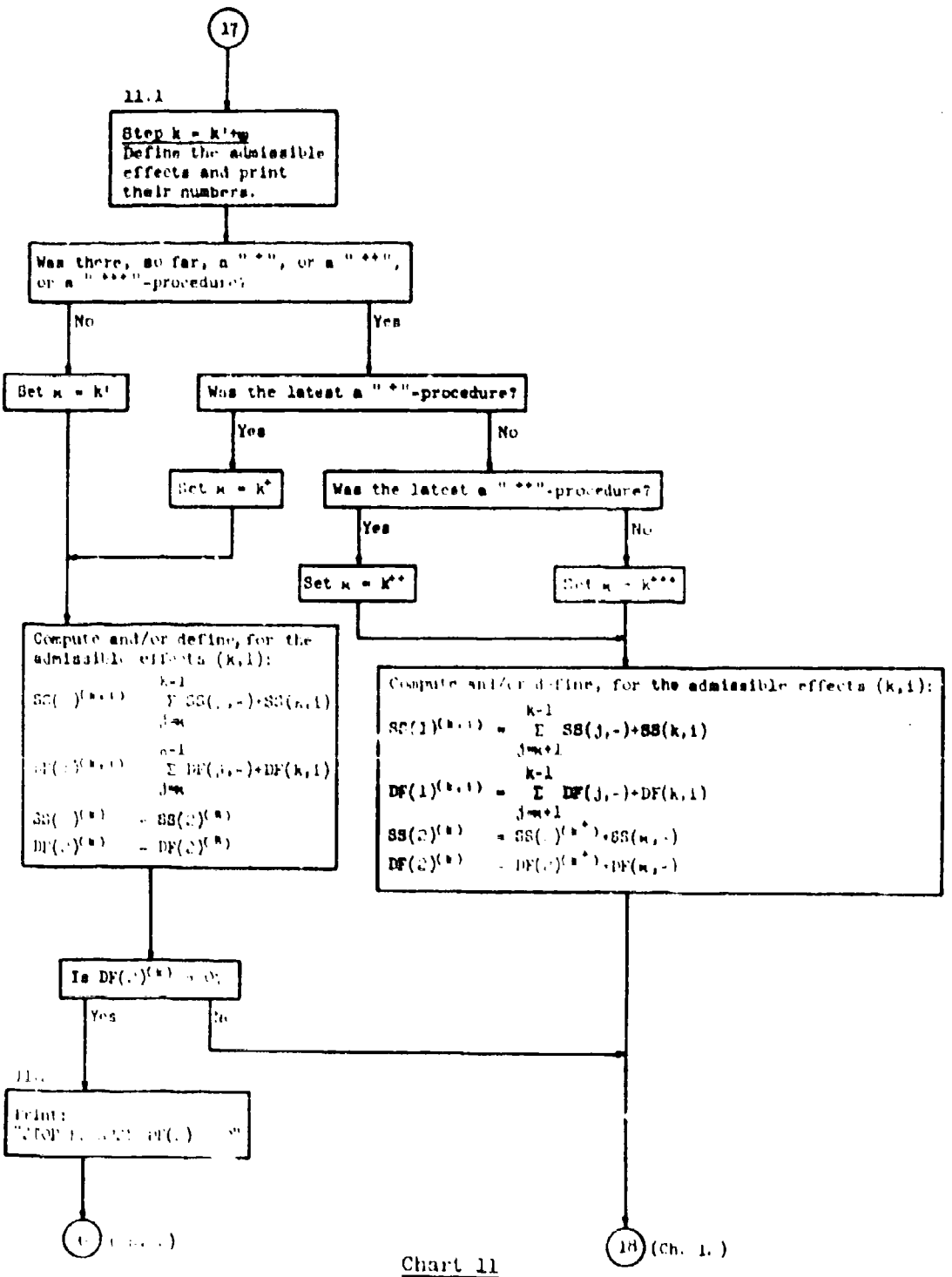

17
11.1
Step k = k'+∞
Define the admissible effects and print their numbers.
Was there, so far, a " + ", or a " ++ ", or a " +++ "-procedure?
No
Yes
Set κ = k'
Was the latest a " + "-procedure?
Yes
No
Set κ = k+
Was the latest a " ++ "-procedure?
Yes
No
Set κ = k++
Set κ = k+++
Compute and/or define, for the admissible effects (k,i):
Compute and/or define, for the admissible effects (k,i):
SS(1)(k,i) = Σ SS(j,-)+SS(k,i)
DF(1)(k,i) = Σ DF(j,-)+DF(k,i)
Is DF(2)(k) = 0?
Yes
Print:
18 (Ch. 1.)

Chart 11

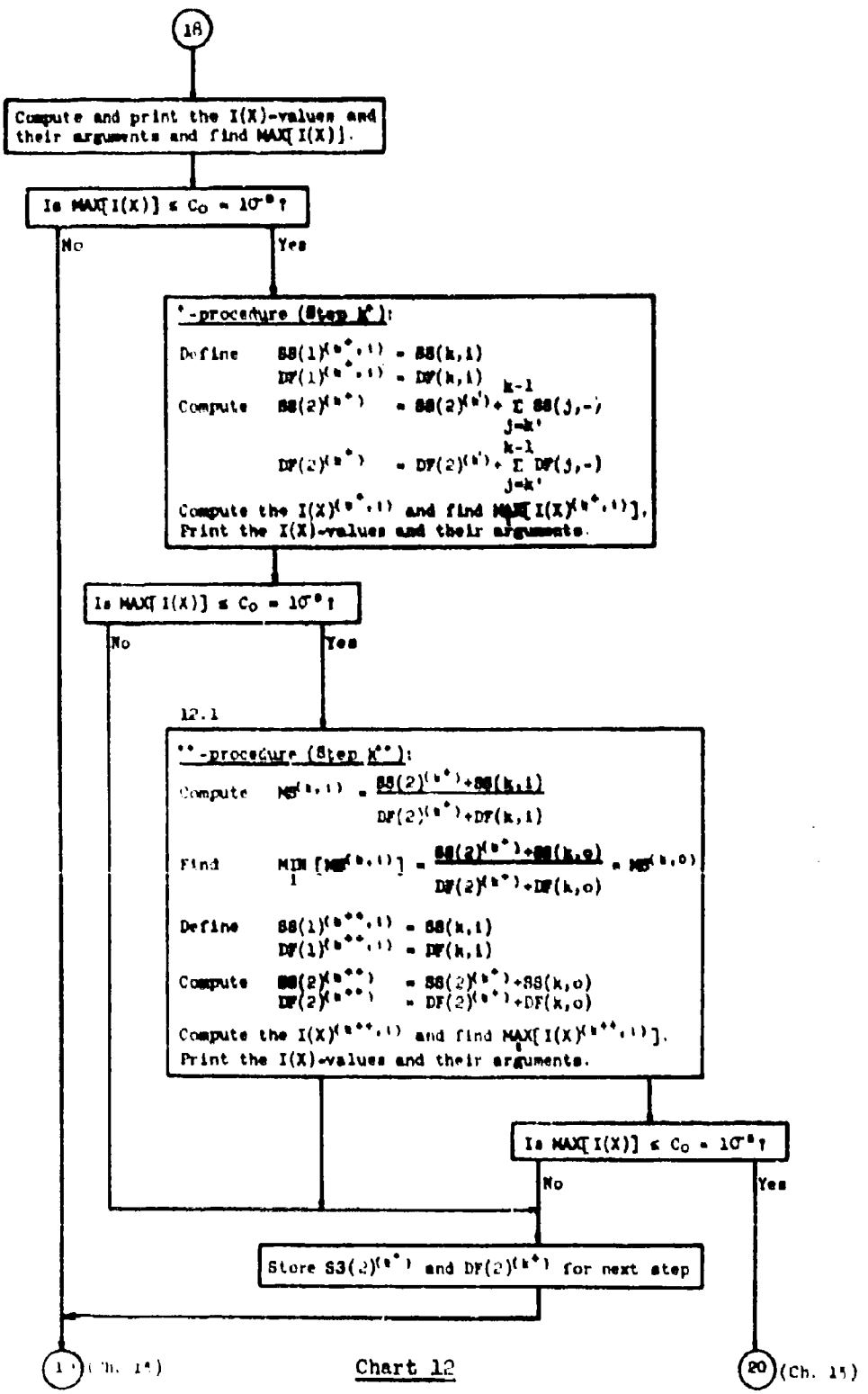

18
Compute and print the I(X)-values and their arguments and find MAX[I(X)].
Is MAX[I(X)] ≤ $C_0 = 10^{-6}$?
No
Yes
-procedure (Step k):
Define $SS(1)^{(k^*,1)} = SS(k,1)$
$DF(1)^{(k^*,1)} = DF(k,1)$
Compute $SS(2)^{(k^*)} = SS(2)^{(k')} + \sum_{j=k'}^{k-1} SS(j,-)$
$DF(2)^{(k^*)} = DF(2)^{(k')} + \sum_{j=k'}^{k-1} DF(j,-)$
Compute the $I(X)^{(k^*,1)}$ and find MAX[$I(X)^{(k^*,1)}$].
Print the I(X)-values and their arguments.
Is MAX[I(X)] ≤ $C_0 = 10^{-6}$?
No
Yes
12.1
-procedure (Step k):
Compute $MS^{(k,1)} = \frac{SS(2)^{(k^*)}+SS(k,1)}{DF(2)^{(k^*)}+DF(k,1)}$
Find $\min_1 [MS^{(k,1)}] = \frac{SS(2)^{(k^*)}+SS(k,o)}{DF(2)^{(k^*)}+DF(k,o)} = MS^{(k,o)}$
Define $SS(1)^{(k^{**},1)} = SS(k,1)$
$DF(1)^{(k^{**},1)} = DF(k,1)$
Compute $SS(2)^{(k^{**})} = SS(2)^{(k^*)}+SS(k,o)$
$DF(2)^{(k^{**})} = DF(2)^{(k^*)}+DF(k,o)$
Compute the $I(X)^{(k^{**},1)}$ and find MAX[$I(X)^{(k^{**},1)}$].
Print the I(X)-values and their arguments.
Is MAX[I(X)] ≤ $C_0 = 10^{-6}$?
No
Yes
Store $S3(2)^{(k^*)}$ and $DF(2)^{(k^*)}$ for next step
1* (Ch. 14)
20 (Ch. 15)

Chart 12

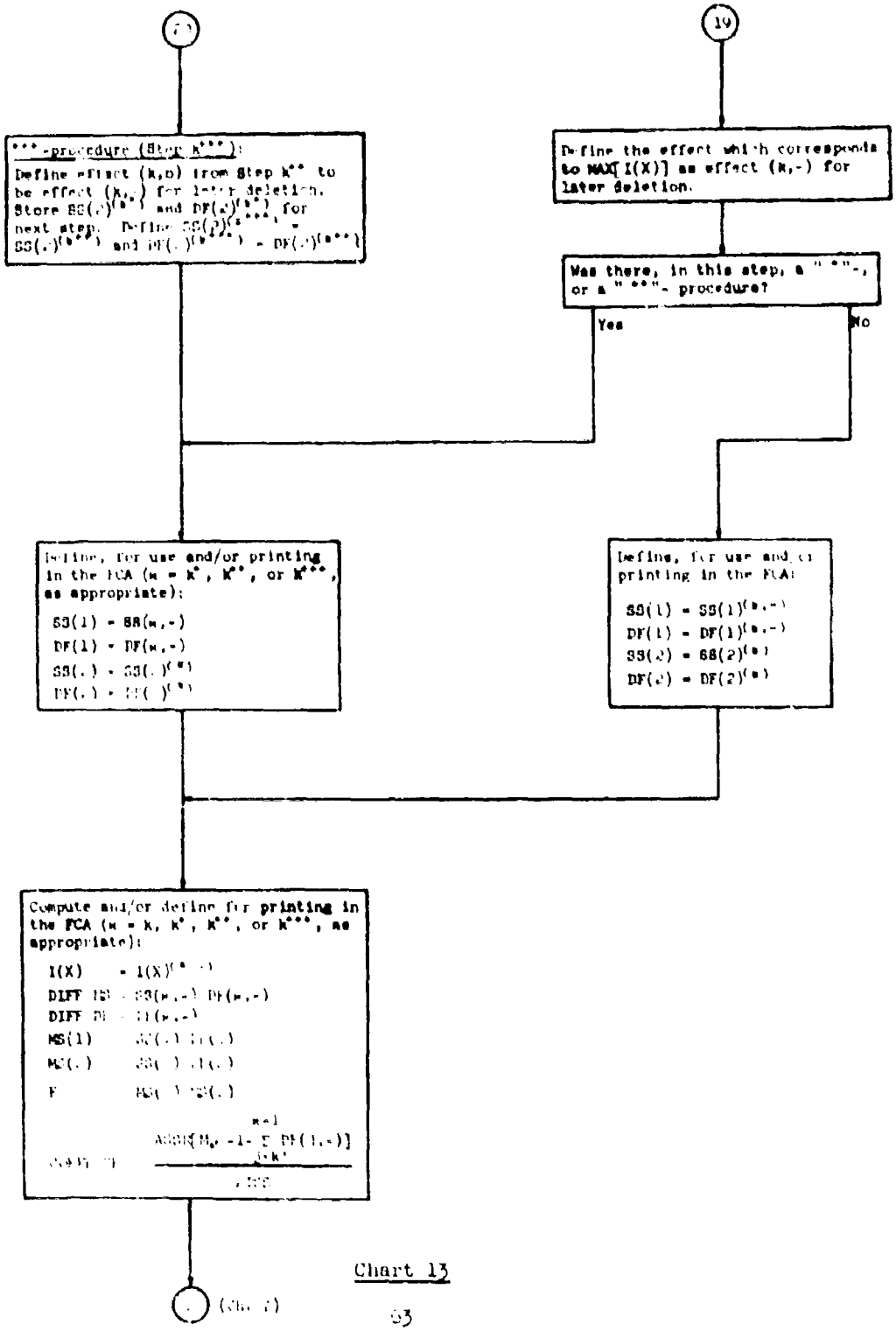

19
Define the effect which corresponds to MAX[I(X)] as effect (H,-) for later deletion.
Was there, in this step, a " ** "-, or a " *** "- procedure?
Yes
No
Define, for use and/or printing in the FCA (H = K*, K**, or K***, as appropriate):
SS(1) = SS(H,-)
DF(1) = DF(H,-)
Define, for use and/or printing in the FCA:
SS(1) = SS(1)
DF(1) = DF(1)
SS(2) = SS(2)
DF(2) = DF(2)
Compute and/or define for printing in the FCA (H = K, K*, K**, or K***, as appropriate):
I(X)
DIFF SS
DIFF DF
MS(1)
MS(.)
F

Chart 13

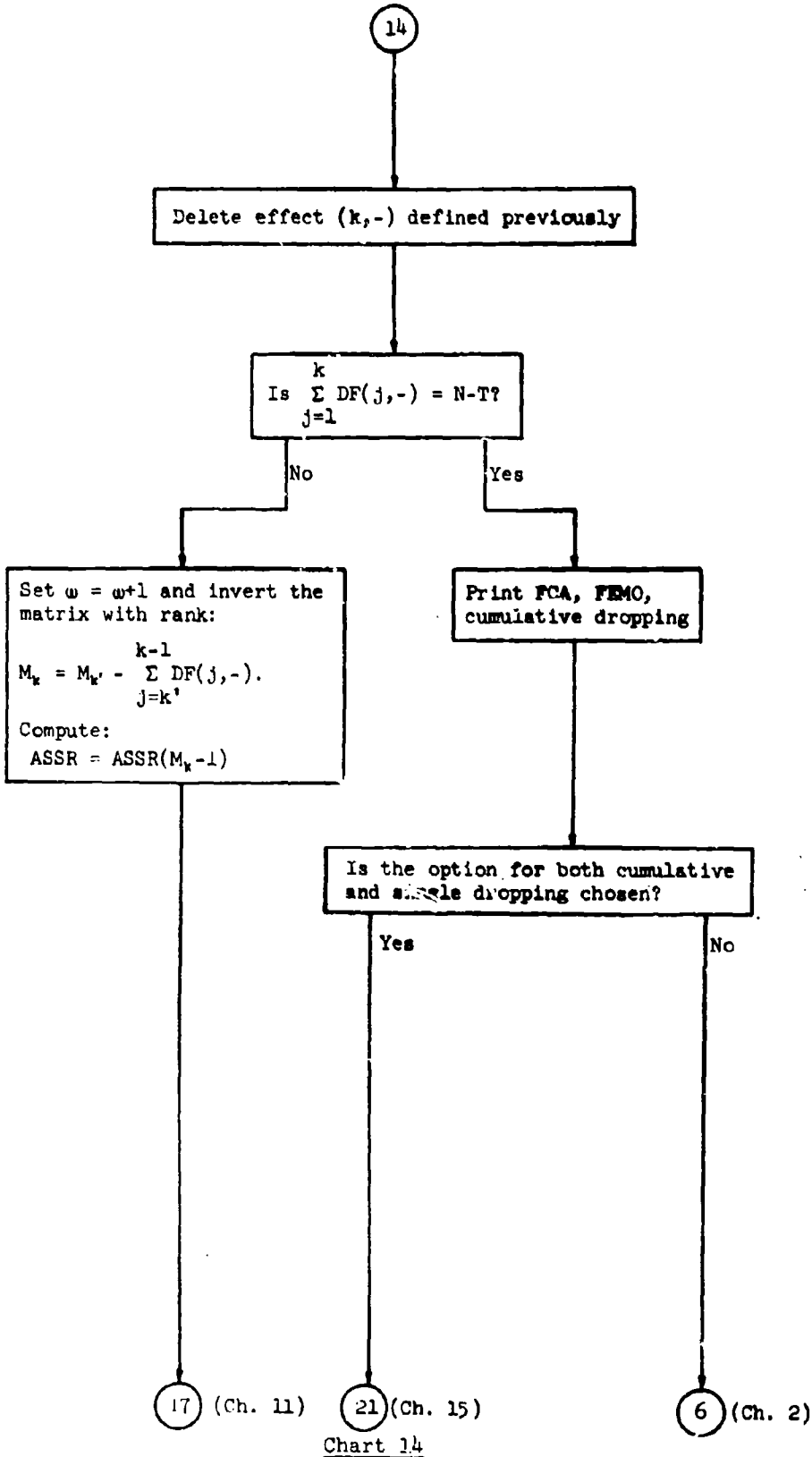

14
Delete effect (k,-) defined previously
k
Is Σ DF(j,-) = N-T?
j=1
No
Yes
Set ω = ω+1 and invert the matrix with rank:
k-1
M_k = M_k' - Σ DF(j,-).
j=k'
Compute:
ASSR = ASSR(M_k-1)
Print FCA, FEMO, cumulative dropping
Is the option for both cumulative and single dropping chosen?
Yes
No
17 (Ch. 11)
21 (Ch. 15)
6 (Ch. 2)

Chart 14

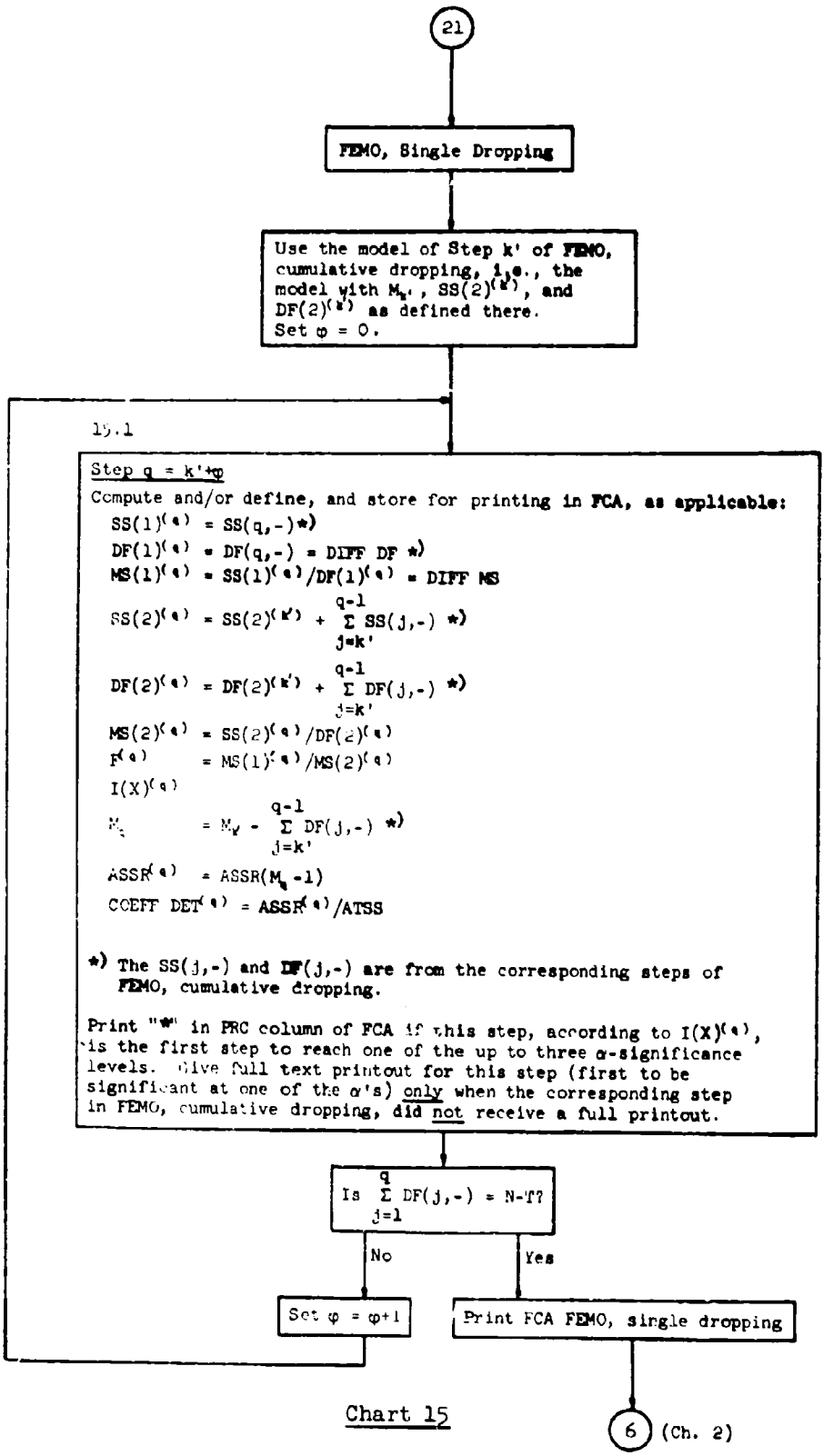

21
FEMO, Single Dropping
Use the model of Step k' of FEMO, cumulative dropping, i.e., the model with $M_{k'}$, $SS(2)^{(k')}$, and $DF(2)^{(k')}$ as defined there.
Set φ = 0.
19.1
Step q = k'+φ
Compute and/or define, and store for printing in FCA, as applicable:
$SS(1)^{(q)} = SS(q,-)$ *)
$DF(1)^{(q)} = DF(q,-)$ = DIFF DF *)
$MS(1)^{(q)} = SS(1)^{(q)}/DF(1)^{(q)}$ = DIFF MS
$SS(2)^{(q)} = SS(2)^{(k')} + \sum_{j=k'}^{q-1} SS(j,-)$ *)
$DF(2)^{(q)} = DF(2)^{(k')} + \sum_{j=k'}^{q-1} DF(j,-)$ *)
$MS(2)^{(q)} = SS(2)^{(q)}/DF(2)^{(q)}$
$F^{(q)} = MS(1)^{(q)}/MS(2)^{(q)}$
$I(X)^{(q)}$
$M_q = M_{k'} - \sum_{j=k'}^{q-1} DF(j,-)$ *)
$ASSR^{(q)} = ASSR(M_q - 1)$
$COEFF\ DET^{(q)} = ASSR^{(q)}/ATSS$
*) The SS(j,-) and DF(j,-) are from the corresponding steps of FEMO, cumulative dropping.
Print "*" in PRC column of FCA if this step, according to $I(X)^{(q)}$, is the first step to reach one of the up to three α-significance levels. Give full text printout for this step (first to be significant at one of the α's) only when the corresponding step in FEMO, cumulative dropping, did not receive a full printout.
Is $\sum_{j=1}^{q} DF(j,-) = N-1$?
No
Yes
Set φ = φ+1
Print FCA FEMO, single dropping
6 (Ch. 2)

Chart 15

2.4.2 Comments on Flowcharts

The following comments refer to the flowcharts and to the boxes in the flowcharts of the previous section, where the flowchart numbers and box numbers are as given.

Chart 1.

Box 1.1 The uncoded OCIV values are printed in order to provide a possibility to check the input. If coding is requested, the ranges and averages (and the range factor C) are printed to facilitate the back-transformation of output values if desired and feasible.

The bar charts (in regression only) of the uncoded OCIV values are printed to give a possibility of visual checks on the distributions of the values, i.e., on their approximate normality and on outlying values.

Box 1.2 See comments Box 1.1.

Box 1.3 The "Full Data Matrix" is the (N+W) x n matrix of the n values (coded, if applicable) of the N IVs and the W dependent variables ($W \leq 4$). It is this matrix from which the (N+1) x (N+1+W) "Summation Matrix" is generated by the program. The summation matrix is the (N+1) x (N+1) matrix of the coefficients of the normal equations augmented by $W \leq 4$ row vectors containing the cross products with the y's. (In most problems, there will be only one dependent variable, y, and the summation matrix will consist of the matrix of the normal equations augmented by one row vector containing the cross products with y.) The ATSS-values (one each for each dependent variable) are the total sums of squares (of y) adjusted for the mean. The ATSS are, naturally, equal for all C. C. 4 sets if there are several such sets.

Chart 2.

Box 2.1 This applies when the problem is one of regression only.

Box 2.2 The additional analyses of variance ("ANVAs") are essentially given in the form of FEMO-FCAs. The ANVAs are computed and printed (1) for each dependent variable for which significant IVs were found in an analysis of covariance and (2) for each OCIV which is significant or is a sub-CIV of a significant GCIV. In addition to the FCAs, the symbols of the admissible effects and all computed I(X)-values and their arguments are printed for each step of each ANVA. See also Sections 2.1.1, 3.3.3 and Example 6 in Section 3.4.6.

Box 2.3 The "Final FCA" is a reprinting of the individual FCAs for each Control Card 4 set. The Final FCA enables the program user to compare more easily the results from the various Control Card 4 sets and, thereby, facilitates the search for the "most probable significant model." See also Section 3.3.2 and Example 5 in Section 3.4.5.

Chart 3.

Box 3.1 Step No. g is a general step of the ranking procedure (COMO or FEMO) before an acceptable inverse of the matrix A of the normal equations is found. See also Section 2.1.3.

Box 3.2 Although the inverse is rejected, A^{-1}, Δ, and the b_v's are printed in order to give the user an impression of the amount of the inaccuracy.

Box 3.3 The same reasoning as given for Box 3.2 applies here. "IDENT. MATRIX DIAG. ELEM. NE 1" is short for "At least one identity matrix diagonal element is not equal to 1, given the tolerance TOLI(2)."

Box 3.4 The value ASSR (Regression sum of squares adjusted for the mean) is given for a rejected first step of an attempted COMO or FEMO because it will enable the user to check the amount of the computational inaccuracy in case of a perfect fit. Namely, in a perfect fit for any data classification into "cells", ASSR should equal the sum of squares "between cells", where the latter can be calculated 'by hand."

Chart 4.

Box 4.1 If there was an attempted COMO without any good steps, g=g-T makes the step numbering in FEMO start with "1."

Box 4.2 The "rightmost" admissible effect to be deleted here is "rightmost" with respect to the position of the DIVs representing the effect in the model. Since the generation of the model is such that the highest-order interactions and the highest powers of the quantitative factor variables are located "rightmost" these are the effects deleted first after FEMO steps have been rejected. (Note that there cannot be statistical criteria by which to delete the effects from the model in this case of a rejected model.) The "minimum degrees of freedom" condition (applicable in FEMO only) serves the purpose of reaching an acceptable model under the smallest loss of degrees of freedom possible. The present way of deleting effects may not be the fastest one to arrive at an acceptable model. In case of a singular matrix A, for example, there may be only one IV which causes the singularity, but this IV may not necessarily be deleted at the first step of an attempted COMO or FEMO. See Example 7 in Section 3.4.7.

Box 4.3 Since there were no accepted FEMO steps, there is no basis for a FEMO, single dropping procedure. (The single dropping procedure takes the ranking order from the cumulative procedure, see Section 2.3.2.)

Box 4.4 The corresponding comments as given for Box 4.2 apply also here, except for the "minimum degrees of freedom" condition since in COMO only individual IVs are deleted.

Box 4.5 See corresponding comments on Box 4.3.

Chart 5.

Box 5.1 h' and k' are the numbers of the "first good step" in COMO and FEMO, respectively. Note that, according to the common matrix A, all steps of FEMO will be accepted when h' was reached in COMO.

Box 5.2 The rank, $M_{h'} = N+2-h'$, of the matrix of the normal equations for the first good step of COMO is the difference of N+1 (for N IVs of the original model and the constant, x_0) and h'-1 (for the h'-1 CIVs deleted prior to the first good step).

Chart 6.

Box 6.1 For more details on the ranking of CIVs by the cumulative dropping procedure of COMO, see Section 2.3.1.

Box 6.2 When DF(2)=0, one deals with a "zero error perfect fit", which can be reached only at the first good step of the ranking procedure. Naturally, MS(2), F, and I(X) have to be defined in this case and cannot be computed. Since $DF(2)=DF(2)^{(h')}$ remains constant throughout COMO, cumulative dropping, the definitions of Box 6.2 apply at each of the remaining steps of COMO, cumulative dropping.

Chart 7.

Box 7.1 If there was a " +", " ++", or " +++"-procedure in any of the previous FEMO steps, this means I(X) had been smaller than $C_0 = 10^{-8}$, that is, smaller than any significance level α specified by the user. Therefore, no full printout will be given (in cumulative or single dropping) beyond this step of FEMO.

Box 7.2 The asterisks printed in the PRC column of the FCA indicate clearly the steps of the ranking procedure where the significant models corresponding to the up to three specified significance levels α have been reached. Note that the asterisk is also printed when, in COMO, a zero error perfect fit was reached at this step. This is because the zero error perfect fit is by definition the first good step, and as such

leads to an I(X) value of zero (also by definition) which necessarily is smaller than any specified α-value. Therefore, only one full printout will be given in case of a zero error perfect fit. This printout is at the same time that of the first good step.

Box 7.3 The option to print the b_ν's at every step of the ranking is provided to supply the user with some information for the intermediate steps where no full printouts are given. The regression coefficients, b_ν, which, for example, in FEMO are the estimates of the individual model parameters, are considered to be the most important numerical values.

Box 7.4 The "full printout" is similar to that given in the program DA-MRCA (Abt et al. [1966]). The Chi-square test computations for testing the normality of the distribution of the residuals are exactly like in DA-MRCA. The "Residual or Error Sum of Squares", the "Check error sum of squares" and the "Square root of (the) residual variance" are specifically computed for the step at which they are printed. That is, the "Residual or Error Sum of Squares" equals ATSS - ASSR(N') when the model of the given step contains N' IVs. In the single dropping procedure, one has ATSS - ASSR(N') = SS(2) where SS(2) is the value which, if divided by DF(2), gives MS(2) as printed in the FCA. (See Section 3.3.1.) A detailed general formulation of the "full printout" in NOVACOM is given in Herring [1967]. See also Section 2.1.4 of the present report and Example 1 in Section 3.4.1.

Chart 8.

Box 8.1 The α-significance level No. KALPHA is the one (specified by the user) which determines the significant CIVs to be kept in the analysis of covariance model when ranking the factorial effects by FEMO. See also Section 2.1.2 and Control Card No. 1, column 25 (Section 3.1.1).

Box 8.2 Since all T CIVs are deleted from the model, FEMO will operate on an analysis of variance model only.

Box 8.3 The significant CIVs will be kept in the model, therefore, FEMO will operate on an analysis of covariance model.

Chart 9.

Box 9.1 The single dropping procedure starts with the model which was that of the first good step of the cumulative dropping procedure.

Box 9.2 For a more detailed description of the single dropping procedure in COMO, see Section 2.3.1.

Box 9.3 The zero error perfect fit is the same as that reached in COMO, cumulative dropping. (See Box 6.2 in Chart 6.) However, since in single dropping the degrees of freedom corresponding to all previously ranked CIVs are accumulated in DF(2), at Step (p+1) one will have DF(2) > 0 and, consequently, MS(2), F, and I(X) can be computed from Step (p+1) on.

Box 9.4 For the case of a perfect fit (zero error perfect fit, that is) see the comments on Box 7.2 in Chart 7.

Chart 10.

Boxes 10.1 and 10.2 See comments on Boxes 8.2 and 8.3, respectively, in Chart 8.

Chart 11.

Box 11.1 For a more detailed description of FEMO, cumulative dropping, see Section 2.3.2. FEMO, cumulative dropping, is presented in Charts 11 through 14 in loop form which is more concise than the detailed manner in which FEMO is described in Section 2.3.2. The various modifications of FEMO, cumulative dropping, (the " +", " ++", and " +++" procedures) in addition to the basic procedure, can be summarized as follows:

No " +" at all: SS(1) and DF(1) are due to the group of all previously ranked (deleted) effects plus the effect presently searched for. SS(2) and DF(2) are those of the first good step of FEMO and remain constant thereafter.

" +": SS(1) and DF(1) are due only to the effect presently searched for. SS(2) and DF(2) are due to the group of all previously ordered effects. At the first step following the " +"-procedure, the pooling starts anew for SS(1) and DF(1), but SS(2) and DF(2) remain constant.

" ++": (This procedure is always preceded by the " +"-procedure.) SS(1) and DF(1) are due only to the effect presently searched for. SS(2) and DF(2) are due to the group of all previously ordered effects plus effect "(k,0)". At the steps following the "++"-procedure, effect (k,0) is replaced, in SS(2) and DF(2), by effect (k^{++},-); and SS(2) and DF(2) remain constant from Step (k^{++}+1) on. Also at Step (k^{++}+1), SS(1) and DF(1) are due only to the effect then searched for. From Step (k^{++}+1) on, pooling starts anew for SS(1) and DF(1).

" +++": (This procedure is always preceded by the " ++"-procedure.) Here, effect "(k,0)" of the " ++"-procedure takes the place of effect "(k^{+++},-)." Otherwise, the " +++"-procedure is as the the " ++"-procedure.

Box 11.2 If DF(2)=0, the data used leads to a zero error perfect fit in FEMO. Since there is no basis, in this case, to rank the factorial

effects by the I(X)-criterion (the computation of I(X) requires an error sum of squares SS(2) > 0), FEMO cannot be executed. The program, therefore, stops and then goes to the next dependent variable or CC 4 set. In order to avoid the stop, the program user must provide for DF(2) > 0 (which will imply SS(2) >0) at the first good step. He may do so by deleting one or more of the factorial effects.

Chart 12.

(See comments given for Box 11.1 in Chart 11.)

Box 12.1 The reason for defining effect "(k,0)" as the one which is to be pooled into the error sum of squares, $SS(2)^{(k^{**})}$, is that (k,0) is the effect which should reasonably be defined as the "least important effect" at this step in case the " ***"-procedure becomes necessary. By previously using (k,0) in the "**"-procedure (one effect has to be defined for pooling into $SS(2)^{(k^{*})}$!), no additional computations are necessary in case of the " ***"-procedure.

Charts 13 and 14.

(See comments given for Box 11.1 in Chart 11.)

Chart 15.

Box 15.1 For a more detailed description of the FEMO, single dropping procedure, see Section 2.3.2. Note that, since in FEMO, single dropping, the ranking order of the effects is taken from the cumulative procedure, there is no need for the " ** " or " ***"-procedure.

3. USE OF NOVACOM

3.1 Input Preparation

3.1.1 Control Cards and Data Cards

In this section the input preparation for NOVACOM is discussed as far as the control cards and data cards are concerned. The consequences of the choices of the various options and the use of the options are described in Sections 3.1.2 and 3.1.3.

The cards are described below in the order of input. The deck of the identification card, the 6 types of control cards and the 3 types of data cards comprise a "problem deck." An arbitrary number of problem decks may be stacked one deck after the other; NOVACOM will perform all the problems in that given order. An end of file card at the end of one problem deck will terminate the NOVACOM computations.

Identification Card

This card contains an 80 column problem identification. The information on this card is completely at the discretion of the user.

Control Card No. 1

Columns	Program Variable	Format	Description	Range
1-2	E	I2	The number of factors. Zero or blank when regression (COMO) only; in which case, columns 4-11 are not used.	0-99
3	NOMODV	I1	The number of dependent variables. Zero, blank or 1 implies one dependent variable.	0-4
4-5	D	I2	The order up to which the program will automatically generate DIVs. Zero or blank when DIVs are to be put in entirely by means of CC No. 4.	0-6 and D ≤ E
6	GDD	A1	GDD = G - generate DIVs described on CC No. 4 and include these DIVs with any which may have been automatically generated, = D - delete the DIVs described on CC No. 4 from those which	G, D, or blank

CC No. 1 (Cont'd)

Columns	Program Variable	Format	Description	Range
			have been automatically generated, GDD = blank - neither generate nor delete DIVs by means of CC No. 4.	
7-9	NCC4	I3	The number of DIVs to be generated or deleted. Blank or zero when GDD = blank. (NCC4 = constant for all NSCC4 sets of CC No. 4, see columns 10-11.)	Generation: 0-139 Deletion: 0-255
10-11	NSCC4	I2	The number of sets of CC No. 4. Zero, blank, or 1 when one set only.	0-99
12-13	TP	I2	The number of OCIVs. Zero or blank when analysis of variance (FEMO) only; in which case, columns 14-19 are not used.	0-99
14-15	P	I2	P = 1 if OCIVs only, or OCIVs and additional hand-generated GCIVs (by means of CC No. 5). $1 < P \leq 6$ if GCIVs to be automatically generated up to powersum P, plus possibly hand-generated or deleted GCIVs (by means of CC No. 5). P = 0 or blank only if OCIVs (TP > 0) to be put in entirely by means of CC No. 5. (An unusual situation.) (Note. GCIVs of powersum 7 to 21 may be hand-generated by means of CC No. 5 only.)	0-6
16	GDC	A1	GDC = G - generate CIVs described on CC No. 5, = D - delete CIVs described on CC No. 5 from the automatically generated set of CIVs, = blank - no CIVs to be generated or deleted by means of CC No. 5.	G, D, blank

CC No. 1 (Cont'd)

Columns	Program Variable	Format	Description	Range
17-19	NCC5	I3	The number of CCs No. 5, i.e., the number of CIVs to be deleted or hand-generated. Blank or zero when CDC blank.	Generation: 0-139 Deletion: 0-254
20	CODE	I1	CODE = 0 - code the OCIVs and the quantitative factor level values. For coding, the form $\frac{x-\bar{x}}{C \cdot R} = x'$ is used, with $\bar{x} = \frac{1}{n}\sum_{1}^{n} x$ and R = max(x) - min(x) and C as specified in columns 21-23 below; n is the total number of observations for OCIV coding or is the number of level values of a factor for quantitative factor level coding. = 1 - do not code.	0,1
21-23	C	F3.0	The coefficient of R in x' (see column 20) for OCIV coding. (C = 1 for quantitative factor level coding.) C = zero or blank has the same effect as C = 1.0.	--
24	DROP	I1	DROP = Zero or blank - cumulative dropping only. = 1 - cumulative and single dropping.	0,1
25	KALPHA	I1	The cardinal number of the ALPHA value which value (in COMO) will be used as the significance level for the inclusion of CIVs in the FEMO model. In case of FEMO only, the program ignores this column.	0-3

CC No. 1 (Cont'd)

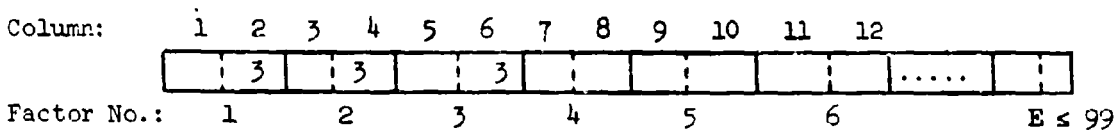
Columns	Program Variable	Format	Description	Range
26-29	ALPHA(1)	F4.4	First significance level for COMO and FEMO	.0001-1.0
30-33	ALPHA(2)	F4.4	Second significance level for COMO and FEMO.	0.0-.9999
34-37	ALPHA(3)	F4.4	Third significance level for COMO and FEMO. NOTE: These values should be in descending order. The program uses only the first non-zero entries.	0.0-.9998
38	CAD	I1	CAD = 0 - use restricted admissibility rules for ranking in COMO and FEMO. > 0 - use unrestricted admissibility rules for ranking in COMO and FEMO (i.e., relaxed admissibility rules for FEMO when both qualitative and quantitative factors are present).	
39-44	TOLI2	E6.2	A tolerance which is used to check the main diagonal elements of the identity matrix formed from the product of the matrix of the normal equations with its inverse. If these diagonal elements deviate from 1 by an absolute value less than the value of TOLI2 then the inverse is considered acceptable.	
45	IRCO	I1	IRCO = 0 - do not print the regression coefficients at every step of NOVACOM. = 1 - print the regression coefficients at every step.	0,1
46	ADA	I1	ADA = 0 - do not perform additional analyses of variance ("ANVAs").	

CC No. 1 (Cont'd)

Columns	Program Variable	Format	Description	Range
			ADA > 0 - perform additional analyses of variance ("ANVAs") for the dependent variable(s) after exclusion of all CIVs from the model and for each significant OCIV and for each OCIV contained in a significant GCIV.	

Control Card No. 2

(Optional - omit when regression only: E = 0.)

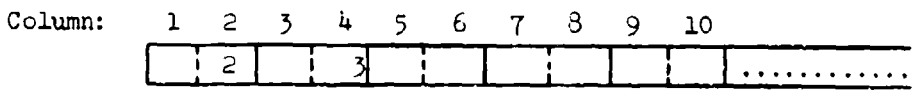


CC No. 2 gives the number of levels of each factor. Example given above: Factors 1, 2, and 3 have 3 levels each.

With two columns per factor there may be entries for 40 factors per card, the maximum number of levels per factor being 99. Use a second card if there are more than 40 factors and also a third card if there are more than 80 factors. Entries are read with an I2 format.

Control Card No. 3

(Optional - omit when regression only: E = 0.)

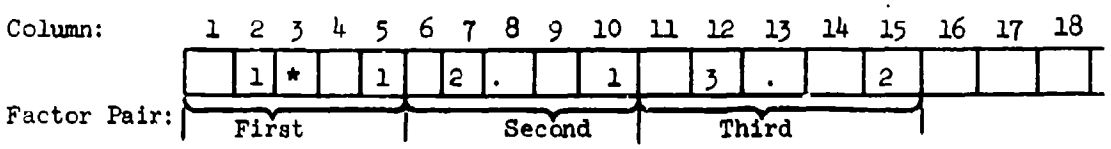


CC No. 3 gives the factor numbers of the quantitative factors using 2 columns for each factor. Example given above: Factors Nos. 2 and 3 quantitative.

Blank card (or cards) when there are no quantitative factors (i.e., all factors are qualitative). Entries are read with an I2 format.

Control Card No. 4

(Optional - omit when E = 0, or GDD = blank and NCC4 = 0.)



Each CC No. 4 gives a DIV to be deleted from the automatically generated set or to be generated (possibly in addition to the automatically generated set of order D). The number of CCs No. 4 equals the number of DIVs to be deleted or generated. (The number of CCs No. 4 is given in columns 7 to 9 of CC No. 1.) There will be 1 DIV symbol per card of CC No. 4 format. Providing 2 digits for each factor number and for each level number or power, the symbol for a DIV of first order will occupy 5 columns; a DIV of second order will occupy 10 columns, etc., (5 columns for each "factor pair"). The maximum order for a DIV is 6. Example above: DIV 1*1 x 2.1 x 3.2.

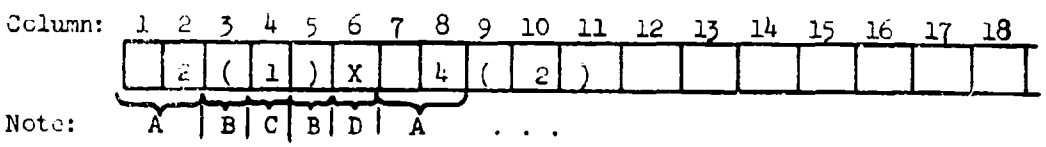
There may be several sets of CC No. 4, each containing the same number of cards = number of DIVs. The number of sets of CC No. 4 is given on CC No. 1, columns 10-11. Each set of CC No. 4 means a separate analysis and, therefore, means a separate final comprehensive analysis (FCA). If there is more than one set of CC No. 4, then the program repeats, as the "Final FCA", all the FCAs for FEMO, cumulative dropping, together as one printout; likewise for FEMO single dropping. Sets of CC No. 4 are stacked one after the other.

When preparing CC No. 4, include in one group DIV descriptions of the same order. Within one of these groups, it is not necessary, but slightly faster, to include together DIV descriptions having the same factors. Arrange the groups by increasing order.

The program works even if the natural (increasing) order of factors in a DIV description is violated. For example, writing either 1*1 x 2.1 or 2.1 x 1*1 is possible.

Control Card No. 5

(Optional - omit if NCC5 = 0 or blank and GDC = blank.)



CC No. 5 (Cont'd)

Note A: two columns for the OCIV number.

Note B: the parentheses are here only for the purpose of conforming with the printout. These columns could be left blank.

Note C: one column for the power.

Note D: X indicates multiplication.

Each CC No 5 describes a CIV to be deleted from the automatically generated set of CIVs or to be generated and included with any CIVs which may have been automatically generated. Hence the number of CCs No. 5 equals the number of CIVs to be deleted or to be "hand-generated" and is given in columns 17-19 of CC No. 1.

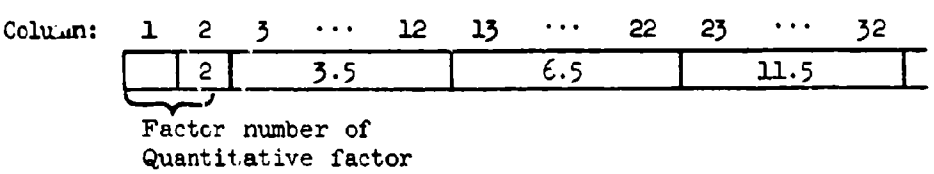
If IP > 0 and no GCIVs are to be generated, P in column 15 of CC No. 1 must be set equal to one.

As an example, GCIV 2(1) x 4(2) ($x_2x_4^2$ in the usual notation) is written in the format illustrated above. In this example the power sum is 3; the maximum power sum for a GCIV is 21. No GCIV may contain more than 6 OCIVs and no OCIV can be raised to a power greater than 9. CC No. 5 input is the only way to obtain GCIVs with power sums > 6.

When preparing CC No. 5, include in one group CIV descriptions with the same power sum. Arrange these groups by increasing power sum. On any CC No. 5 the OCIV numbers must be in increasing order from left to right.

Control Card No. 6

(Optional - omit when there are no quantitative factors.)



A set of CC No. 6 gives the (uncoded) quantitative factor levels for one of the factors which are indicated as being quantitative on CC No. 3. These sets of CC No. 6 should be in the same order as the quantitative factor numbers are entered on CC No. 3.

CC No. 6 (Cont'd)

In columns 1-2 of the first of a set of CC No. 6 for a particular factor there must be punched, right adjusted, the factor number. In columns 3-12 the value of the first level is entered, in columns 13-22 the value of the second level, ..., in columns 63-72 the value of the seventh level. The value of factor level 8 would begin on the next card in columns 3-12 (with columns 1-2 blank) and so on until the values of all levels have been entered for this factor. The level values are read with an F10.5 format.

As an example, the values of the three levels of quantitative factor number 2 are entered: 3.5, 6.5, and 11.5 for the first, second, and third level, respectively.

Data Cards

1. The 1st data card (optional - omit if regression only: E = 0) gives the cell identification, using two columns for each factor and as many cards as necessary, until a level number has been entered for each factor. If there are several observations of the dependent variable(s) y in one cell, the first data card(s) must be repeated for each of them.

Example: Cell 131

Column:	1	2	3	4	5	6	7	8
		1		3		1		
Factor No.:	1		2		3		4	

2. The 2nd data card gives the value(s) of the dependent variable(s) y in the cell identified by the 1st data card and/or associated with the CCIV values entered on the 3rd data card(s). Each y-value occupies 10 columns beginning with columns 11-20. The values of up to 4 dependent variables may be entered depending upon the value of NOSODV in column 3 of CC No. 1. In order to specify the last "2nd data card", columns 1-4 of the 2nd data card must have the characters LAST. Otherwise, columns 1-10 are blank. The y-values are each read with an F10.5 format.

Example: NOSODV=2; this card is the last 2nd data card in the problem deck; the value of the first dependent variable is 2.0 and the value of the second dependent variable is 4.5.

Column:	1	2	3	4	5	6	7	8	9	10	11 ··· 20	21 ··· 30
	L	A	S	T							2.0	4.5

3. The 3rd data card (optional - omit when there are no OCIVs: TP = 0) gives the (uncoded) values of the OCIVs as they are observed together with the y-values(s) entered on the 2nd data card. The number TP of OCIVs is given in column 12-13 of CC No. 1. Each OCIV value occupies 10 columns, 8 values per card. Each OCIV value is read with an F10.5 format. If there are several observations of the dependent variable(s) y for a given set of OCIV values, the third data card(s) must be repeated for each observation.

3.1.2 Model Generation Options

As can be seen from the description of the Control Cards, the generation of the N IVs of the NOVACOM model is controlled by the entries in columns 1 to 19 (except column 3) of CC No. 1 and by the entries in Control Cards No. 2 to 5. The options for the generation of the FEMO part of the model will be discussed first.

The data classification to be analyzed by NOVACOM may have a maximum number of E=99 factors (column 1, CC No. 1). Since the limitation of the number N of IVs in the model is 139, the restriction on the feasible number of factors will be severe in most cases.

The order D (columns 4-5 of CC No. 1) up to which the program will automatically generate DIVs, cannot be larger than 6. This means that the highest order interaction which can be included in the FEMO part of the model is that among 6 factors. (The maximum order 6 of DIVs is also reflected by the specifications of CC No. 4.) For example, if the user has a case with E=6 factors, f_1, f_2, ..., f_6, and specifies D=6, the program will automatically generate $\prod_1^6 F_j - 1$ DIVs, where F_j is the number of levels of factor f_j, $j = 1,2,\ldots,6$. With $F_j = 2$, that is, with a 2^6 - data classification, $2^6 - 1 = 63$ DIVs will be automatically generated. The upper limit of the number of automatically generated IVs (DIVs and/or CIVs) is 255. For example, the DIVs of a four-factor classification, where each factor has 4 levels, could be generated automatically by specifying E=D=4: There are $4^4 - 1 = 255$ DIVs in the full model of this case. However, since the model limitation is N=139, at least 255 - 139 = 116 DIVs would have to be deleted from the automatically generated set of 255 DIVs. The deletion, in this case, would have to be done via CC No. 4 by specifying the 116 DIVs to be deleted.

The program variable GDD, column 6 of CC No. 1, controls the 4 options of the generation of the FEMO part of the model; see further below.

The number NCC4 of Control Cards 4 (columns 7-9 of CC No. 1) equals the number of DIVs to be deleted or "hand"-generated. The range of

NCC4 indicates that it is theoretically possible to generate up to 139 DIVs by hand (i.e., to go to the limit of N=139 IVs), or to delete up to 255 DIVs from the automatically generated set of up to 255 DIVs. In general, the user will generate or delete considerably fewer DIVs by means of CC No. 4.

When the number NSCC4 (columns 10-11, CC No. 1) of sets of CC No. 4 is larger than 1, the number of DIVs = number of cards, NCC4, is the same for each of these sets. The only reason for having two or more sets of CC No. 4 is the presence of two or more, respectively, confounded constants for the data layout, see Appendix A. When there are two confounded constants, it means that 2 models can be fitted which differ only in one DIV, that is, N-T-1 DIVs are identical in both models. The one remaining DIV represents, in each "possible" model, a different one-degree-of-freedom-effect, where these two effects are completely confounded. If more than two constants are confounded for a data layout, there are more than two "possible" models which are represented by more than two sets of CC No. 4. FEMO executes the ranking of the factorial effects for all NSCC4 models put in by the corresponding number of sets of CC No. 4. See also Section 3.3.2 and Example No. 5 in Section 3.4.5.

Obviously, there is a number of possibilities to generate the FEMO part of the NOVACOM model. The set of the N-T DIVs (with possibly T=0) of the model may be generated in the following 4 ways:

I. by automatic generation only;

II. by automatic generation and "hand"-generation via CC No. 4;

III. by automatic generation and deletion via CC No. 4;

IV. by "hand"-generation only.

I. "Automatic generation only" (by the program variable D, columns 4-5, CC No. 1) is applicable only when a "full model" is to be generated. For the ANOVA part in FEMO, a "full model" means that all factorial effects up to a given order (D) are to be generated and can be generated which requires the presence of observations in all cells of the associated data classifications. For example, in a three-way classification with factors $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, all DIVs representing main effects and 2-factor interactions may be generated (D=2) if none of the three two-way classification tables has empty cells: the $\mathcal{A} \times \mathcal{B}$, the $\mathcal{A} \times \mathcal{C}$, and the $\mathcal{B} \times \mathcal{C}$ table. (Note. In the above discussion the absence of "identities" was assumed, see Appendix A.)

II. "Automatic generation and hand-generation" of the DIVs may be used in order to save on input writing in cases where a "full model" is not to be generated. For instance, in the example mentioned under "I" above, the user may wish to fit and be able to fit some DIVs representing degrees of freedom of the three-factor interaction in addition to the full model of order D=2. Rather than writing all N-T DIVs of the model on CC No. 4 (which the user may do if he wishes to, see "IV." below), the user may automatically generate the second order model (D=2) and write only the additional third order DIVs by means of CC 4. See example 4 in Section 3.4.4.

III. "Automatic generation and hand-deletion." Again taking the above example, this third way of model generation enables the user to automatically generate the third-order model (D=3) and then to write on CC No. 4 the DIVs representing those degrees of freedom of the three factor interaction which are not wanted in the fit (or cannot be fitted) and are to be deleted. (See Example 5 in Section 3.4.5.) The method of input (II or III) in such a case is left to the user. In general, the user will choose the way which means the least amount of input writing via CC No. 4.

IV. "Hand-generation only." This option may be useful in some cases when the input writing of DIVs to be deleted from an automatically generated set involves more work (and more possibilities of writing errors!) than would be encountered in specifying the whole set of N-T DIVs, to be generated, on CC No. 4. See Example 6 in Section 3.4.6.

The options for the generation of the COMO part of the NOVACOM model are very similar to those of the FEMO part.

The program variable P, columns 14-15, CC No. 1, is the "power-sum" up to which the program will automatically generate CIVs. The power-sum is defined, as the name suggests, as the sum of all powers in a CIV. The GCIV $x_1^2x_3^3x_4$, for example, has a power sum of 2+3+1=6.

The power-sum, P, up to which the program will automatically generate CIVs, is equivalent to the "order" of the CIV-model as it was called in Section 2.2.1. The reader is referred to that section and to formula (2-14) giving the total number, T, of CIVs in the model when P is specified. A "full model" of order P means that all GCIVs may be generated which, in general, is the case if no linear dependencies are introduced into the matrix A by this generation. For a more detailed discussion of linear dependencies ("obvious" and "non-obvious") in regression models see Reference 2.

While the upper limit for P is 6, GCIVs of higher order, or larger power sum, may be "hand"-generated by means of CC No. 5. The maximum power sum of GCIVs thus generated is 21.

The options I - IV described before for the generation of DIVs apply correspondingly for the generation of CIVs. The 4 options are controlled by the program variables GDC (column 16) and NCC5 (columns 17-19) of CC No. 1. As a consequence of the 4 options to generate each part of the NOVACOM model, there are, in a model which is to contain both CIVs and DIVs (i.e., in an analysis of covariance model), 4x4=16 different ways of generating the same model, provided one deals with "full models" in both DIVs and CIVs.

3.1.3 Ranking Options

The ranking options in COMO and FEMO are controlled by the entries in columns 24 to 38 of CC No. 1.

The input value for the program variable DROP, column 24, determines whether cumulative dropping alone is performed in COMO and/or FEMO (DROP = 0), or both cumulative and single dropping (DROP = 1). Since in single dropping the ranking order of CIVs or factorial effects is taken from the order established with the cumulative dropping, the additional running time for single dropping is small. Therefore, the program user will probably choose, in most cases, the option for both cumulative and single dropping.

The difference between the two dropping procedures is in the determination of the significant model. Since the two ranking orders are identical, the difference between the procedures is only the step at which the "non-significance," i.e., the $\Sigma(X)$-value, reaches a given significance level α (see columns 26-37 of CC No. 1). In some cases, the two significant models will be identical; in other cases, they will be different. Therefore, the user who chooses the option for both dropping procedures may face the problem of having to decide between two significant models.

The problem is similar to the one encountered in "orthogonal" ANOVA: Should one pool the non-significant effects into the error term or not? Whereas in orthogonal ANOVA the reason for pooling is usually the desire to increase the number of the degrees of freedom for error, the pooling in NOVACOM is a feature of the ranking method employed here. There is pooling in both dropping procedures. In cumulative dropping, the pooling takes place in the numerator mean square of the F-value (of the Main Theorem, see (2-10) in Section 2.1.2), whereas in single dropping the pooling takes place in the denominator mean square. Among the two ranking procedures of NOVACOM the cumulative dropping procedure obviously is the "right" one, because single dropping implies a redefinition of the model at each step according to an intermediate result of the analysis. However, the single dropping is provided as an additional procedure since cumulative dropping tends to be less powerful than single dropping. An example may serve as an illustration: In the ranking order, as established by FEMO, cumulative dropping, of the factorial effects in a given problem, the first k-1 effects ranked as least important may, in the true model of

the problem, be non-existent. Their associated mean square is then an estimate of the error variance. The kth effect, however, may exist and may have a relatively large contribution to the numerator mean square in the F_c-value of the Main Theorem. But as a consequence of the pooling with the k-1 non-existing effects, the I(X)-value at Step k in FEMO may only be slightly decreased as compared to Step k-1. In contrast to this, single dropping would "detect" the significance of the kth least important effect since it would have pooled all the k-1 non-existing effects into the error term where they actually belong according to the assumption made for this example: the I(X)-value at Step No. k of single dropping would definitely be much smaller than that of Step No. k-1, and would indeed possibly reach a given significance level α.

In many cases the gap between the two significant models may be closed by applying essentially the equivalent of the alternative ranking procedure discussed in Section 2.1.2, provided one makes the (reasonable) assumption that the ranking orders resulting from the cumulative and the alternative ranking procedures would be identical. In that case, the non-significant I(X)-values at the steps of the cumulative dropping procedure indicate which null hypotheses may be accepted. At each such step where all previous null hypotheses could be accepted, the additional regression sum of squares due only to the effect ranked at that step may be divided by the associated degrees of freedom to give a mean square ("DIFF MS" in the FCA; see Section 3.3.1) which has expectation σ^2 if the null hypothesis concerning the effect ranked at the step is true. Therefore, if in the single procedure the I(X)-value reaches a given significance level α^* at a step No. k_1, say, and if, in the cumulative dropping procedure, I(X) reaches α^* at a later step No. k_2, say, $(k_2 > k_1)$, one can divide DIFF MS from Step k_1 in the cumulative procedure by the original error mean square based on n-N-1 degrees of freedom to form a valid F test. If this F-value is significant at the same level as the F-value at Step k_1 in the single dropping procedure, the two significant models are identical and the gap is closed. See also Section 3.3.1 and the discussion of the numerical examples in Section 3.4.

The user of NOVACOM should keep in mind that the main purpose of the program is to screen incomplete and unbalanced data classifications for significant factorial effects. NOVACOM, by its nature, cannot always give clear-cut answers such as may be possible in "orthogonal" analysis of variance.

Therefore, if the gap between the two significant models cannot be closed by any means, the statistician may conclude that the true significant model is between the two models from the two procedures. This situation will indicate the need for additional efforts to further analyze the given body of data. (See Example 6 in Section 3.4.6.)

The variable KALPHA (column 35 in CC No. 1) specifies which of the possibly up to three significance levels α (columns 20 to 37) will be used for the determination of the significant model in COMO, cumulative dropping, when the model also contains a FEMO part. That is, KALPHA has importance only in the case of an analysis of covariance. (In case of regression only the specified value of KALPHA has no influence upon the printout.)

The variable KALPHA would not have been required in NOVACOM (for analysis of covariance cases) assuming there was only one α-value (instead of three) to be specified for the determination of the significant model. However, for reasons given below, up to three such α-levels may be specified by the user and, therefore, one of them has to be chosen for COMO by means of the variable provided, KALPHA. (The program could have been constructed such that up to three FEMOs would have been performed corresponding to the three α-values; however, this possibility was disregarded in order not to add unnecessarily to the computer running time of a given problem.)

Since KALPHA determines which one of the three significance levels is to be used in COMO, cumulative dropping, for FEMO, and considering the fact that cumulative dropping in some cases yields significant models which contain fewer terms than actually are significant, the choice of the three α-values and KALPHA should be made accordingly. That is, in analysis of covariance it will be advantageous to choose the first of the three α-values larger than actually desired for the determination of the significant model and then set KALPHA=1. For example, if $\alpha = 0.01$ is the desired level for the significant model in an analysis of covariance case, one could choose $\alpha_1 = 0.05$, $\alpha_2 = 0.01$, and $\alpha_3 = 0.001$, say, where then, with KALPHA=1, the significant CIVs to be carried through the FEMO part of the ranking would be determined at the 0.05 level of significance in COMO, cumulative dropping. See also Example 6 in Section 3.4.6.

The above is one reason for having the possibility to specify more than one α-level of significance in NOVACOM. Another reason is that the program gives a "full printout" containing all pertinent information for a given step of the ranking only when a specified α-level is reached by I(X). By specifying the maximum of 3 α-levels, say, NOVACOM will give, at the most, 3 full printouts for both cumulative and single dropping in both COMO and FEMO (other than for the first good step), provided the 3 levels are reached by I(X) at different steps in each procedure. (Note. NOVACOM could have been constructed such that the "full printout" would have been given at each step of the rankings, however, this possibility was disregarded because of problem running time considerations.)

The analyst, in choosing more than one significance level and in obtaining the corresponding full printouts, is enabled to broaden his judgment concerning the significance of the CIVs and/or factorial

effects. By choosing two "neighboring" significance levels in addition to the one principal level decided upon in advance, the user obtains all information about the models which would have resulted from the other two significance levels. In general, the user will choose, if he makes use of the option, one α-level above and one below the one principal significance level. For example, he may have chosen $\alpha = 0.05$ as his principal significance level. He would write $\alpha = 0.05$ into columns 30-33 of CC No. 1 as ALPHA(2) if he would like to specify two additional values. These could be ALPHA(1) = 0.10 and ALPHA(3) = 0.01, for example. It should be noted that the α-values must be put in by decreasing order, that is, ALPHA(1) > ALPHA(2) > ALPHA(3). This order is in accordance with the characteristics of the backward ranking technique.

The variable CAD (column 38 of CC No. 1) specifies whether to use restricted or unrestricted (relaxed) admissibility rules in the ranking process. The terms "restricted" and "unrestricted" concern only the admissiblity rules for ranking CIVs and/or factorial effects containing quantitative factors. (The ranking of factorial effects with respect to the qualitative factors contained in them is always done under "restricted" admissibility rules.) The choice of the type of admissibility rules is entirely up to the user. Restricted admissibility in the ranking of CIVs or factorial effects containing quantitative factors is necessary if the option for coding the OCIV values and level values is chosen (in order to achieve computational accuracy) provided the user wishes to have a possibility to retransform the resulting values (for example, the regression coefficients) without changing the significant model, see Abt et al. [1966]. Another application of restricted admissibility is when a significant model is desired which is to contain all polynomial terms having lower order than the significant terms. For example, a significant model may result, under unrestricted admissibility rules, which contains only second order terms. If the user wishes, in this case, for reasons of physical interpretation, a model also containing the first order (linear) terms, he can achieve this by applying the restricted admissibility rules.

Ranking under restricted admissibility rules is also the only means to arrive at a breakdown of the sums of squares which corresponds to the breakdown achieved by the method of orthogonal polynomials. As known, orthogonal polynomials are constructed such that each polynomial term is fitted "in addition" to all previously fitted terms. This holds independently of the fact whether one uses orthogonal coefficients in the case of equidistant levels or actually constructs the polynomials in the case of non-equidistant levels. Therefore, in the use of orthogonal polynomials, the quadretic contrast as such, for example, is meaningless; only the fact that it is fitted in addition to the linear contrast gives it meaning. This feature of fitting "in addition" to all lower order terms is achieved by the restricted admissibility rules in NOVACOM.

If none of the three discussed reasons for using restricted admissibility rules in the ranking are present, the analyst should choose the option for "unrestricted admissibility" (CAD=1). He may even have a strong reason to do so because the relationship he is analyzing statistically may be theoretically known to contain no linear terms, for example. One can often observe that ranking (by NOVACOM) of CIVs or effects under "unrestricted admissibility" leads to significant models containing many fewer terms than result from ranking under restricted admissibility. (See Example 3 in Section 3.4.3.) That is, unrestricted admissibility may lead to a significant model in which a minimum number of terms can explain a maximum of the total variability.

3.2 Running Time Formula

The running time needed by NOVACOM for a given problem obviously is dependent upon many parameters. In order to find an approximate time formula, a prediction equation was evaluated by applying the program DA-MRCA to the actual running times of a number of problems which had been run with NOVACOM on the IBM 7030 (STRETCH). In this study, the actual running time used by NOVACOM was the "dependent variable", y, in minutes. As "independent variables" (OCIVs, that is) the following three variables were used: x_1 = N = number of IVs; x_2 = G = number of factorial effects; and x_3 = n = number of observations. The BIVOR-subroutine of DA-MRCA led to a more concise formula as compared to the one resulting from IVOR. In order to account for the time consumption caused by some of the other parameters, the coefficients W, S, and H (see below) are introduced in the formula. All actual running times which entered the least squares evaluations are from problems where the ranking option for both cumulative and single dropping was chosen. In previous studies, only little differences were noted between the running times of rankings with restricted and unrestricted admissibility rules. The parameter restricted/unrestricted admissibility is, therefore, neglected in the formula. The number of full printouts in the problems, whose times were used in the evaluation, may be considered as representative of the typical problem.

The formula is as follows (T = NOVACOM - time in minutes on IBM 7030 STRETCH):

$$T = 0.6 + \frac{WS}{10^6}\left[114nN + 20HGN^2 + 0.159Gn^2\right]$$

where the symbols have the following meaning:

W = number of dependent variables

S = number of sets of CC No. 4

H = number of ANVAs performed (this must be estimated since H is dependent upon the results of the analysis)

n = number of observations

N = number of IVs

G = number of effects. (In case of multiple regression, use G=N.)

In the use of the formula, the third term in the expression in brackets may be neglected as long as $n < 100$, say. Only in analysis of covariance has H to be estimated. In the two other types of problems, H=1. $S > 1$ applies only in cases with confounded constants; otherwise, S=1.

Since the estimated standard deviation for the least squares fit of BIVOR was 0.9 (minutes), the formula is not very precise for the running times of small problems. (See also the running times printed for the examples in Section 3.4.) For example, for G=10, N=20, and n=100 (W=S=H=1, say), the formula yields:

$$T = 0.6 + 10^{-6}[114 \cdot 100 \cdot 20 + 26 \cdot 10 \cdot 400 + 0.159 \cdot 10 \cdot 10000]$$

$$= 0.6 + 10^{-6}[228000 + 104000 + 15900]$$

$$= 0.6 + 0.348 = 0.948 \approx 1 \text{(minute)}.$$

However, the actual running time for a case like this may be as much as three minutes. The relative accuracy is much better for large cases for which the formula is mainly intended. In the largest case used for the time study, where n=768, N=138, G=88, the actual running time of NOVACOM was 6[illegible] minutes. The predicted time, by the formula given, for this case is 64.9 minutes.

3.3 Interpretation of Results

In this section the meaning and use of the results contained in the Final Comprehensive Analysis printouts of NOVACOM will be discussed. The formulation of the complete printout for a problem is not given in algebraical terms. However, the complete printout is discussed with Example No. 1 in Section 3.4.1. For a general formulation of the complete printout see Herring [1967] and the general interpretation of the printout of DA-MRCA (Reference 2) which is similar to the complete printout of NOVACOM. (Note. "Complete printout" means the entire printout for a problem; whereas the "full printout" consists of the pertinent data at a significant step in the ranking.)

3.3.1 The Final Comprehensive Analysis (FCA)

The format of the Final Comprehensive Analysis is the same for COMO, FEMO, and for the ANVAs. (In case of more than one set of

CC No. 4, the FCAs of FFMO are merely repeated as the "Final FCA" at the end of a problem printout.) If the corresponding option is requested (column 24 of CC No. 1), the FCA is printed for both the cumulative and the single dropping procedure in COMO and/or FEMO.

The FCA format has 12 columns which are discussed below, starting from the left.

"STEP". This column gives the step number of the ranking procedure. (These are the same numbers which identify the "full printouts.") If COMO has been part of the ranking, FEMO will always start again with step number "1." The step numbers are not influenced by the fact that one or more models could not be accepted by the program. (See Example 7 in Section 3.4.7.)

"CONC VAR" (COMO) or "EFFECT" (FEMO). The second column from the left gives the symbol of the concomitant variable ranked at a given step of COMO or the symbol of the effect ranked at a given step of FEMO. The symbols used are explained in Section 2.2.

"PRC". In this column the PRoCedure is indicated which was used in the ranking of an effect at a given step of FEMO. Depending upon whether the $^{+}$-, $^{++}$-, or $^{+++}$-procedure occurred at the given step, the corresponding symbol is printed in this column. Also printed in the PRC column is an asterisk if, at the given step, I(X) reached one of the up to three specified significance levels α for the first time. Since even the smallest specified α-value will be larger than the value $C_0 = 1\sigma^8$ which, if reached by I(X), activates the $^{+}$- (or $^{++}$-, or $^{+++}$-) procedure, the printing of the symbol " $^{+}$" (or " $^{++}$", or " $^{+++}$") always has predominance over the asterisk.

An asterisk indicates which step in the ranking corresponds to the significant model for the α-level which is associated with that asterisk. That is, the CIVs or effects whose symbols are printed at the step number where the asterisk occurs and at all higher step numbers belong to the significant model corresponding to that asterisk.

"I(X)". This column gives the computed value of I(X) which is associated with the CIV or effect ranked at a given step. In general, the printed I(X)-values will decrease with increasing step numbers. Due to the behavior of the values which enter the I(X) computation, however, the I(X)-values may fluctuate considerably in some cases.

Naturally, the asterisk in the PRC column at a given step corresponds to the I(X)-value which reaches, for the first time, a value smaller than or equal to the significance level α associated with that asterisk. For example, if the first asterisk printed corresponds to $\alpha = 0.05$, the I(X)-value of this step will be smaller than or equal to 0.05.

"DIFF MS". The abbreviation of this column stands for "DIFFerence Mean Square." The value printed is the difference between the regression sums of squares of two consecutive steps in the ranking, divided by the degrees of freedom of the CIV (which is always 1) or the effect ranked at the given step. In "orthogonal" analysis of variance, DIFF MS equals the mean square which one would obtain in a regular ANOVA table (see Examples 2 and 3B in Section 3.4.).

In "non-orthogonal" ANOVA, as was discussed in Section 3.1.3, DIFF MS can be used as a basis for a valid F test only when the null hypotheses corresponding to all previously ranked CIVs or effects could be accepted. The user would, in this case, divide DIFF MS by MS(2) (given in a column of the FCA discussed further below) to obtain an F-value whose significance he can find from an F-table.

In the single dropping procedure where the user is willing to redefine his model at each step of the ranking order (which was established by the cumulative ranking procedure), DIFF MS actually is the basis of the F-value printed in an FCA column more to the right.

"DIFF DF". The abbreviation of this column stands for "DIFFerence Degrees of Freedom," and the number DIFF DF printed is associated with DIFF MS as indicated before. Independently of that association, the DIFF DF column takes the place of the usual degrees-of-freedom column in a "regular" ANOVA table.

"F". This column gives the F-value of the Main Theorem, see (2-10) in Section 2.1.2, as computed for the CIV or effect ranked at the given step, for the cumulative or the single dropping procedure, as applicable.

"MS(1)". This column gives the computed mean square of the numerator in F of the previous column. In single dropping, MS(1) equals DIFF MS for obvious reasons.

"DF(1)". This is the number of degrees of freedom in the numerator of the value in the F-column. If, in cumulative dropping, DF(1) is multiplied by the MS(1)-value of the previous column, the result is the "additional regression sum of squares", $SS_{N-N'}$, of the Main Theorem, due to the N-N' IVs which have been deleted at the given step (when the number of IVs remaining in the model is N'). In single dropping, one has merely DF(1) = DIFF DF and DF(1) · MS(1) = DIFF MS.

"MS(2)". This column gives the computed mean square of the denominator in the F-value printed in the F-column. MS(2) is the estimate of the residual variance and is used as such in the ranking procedure. In cumulative dropping, MS(2) remains constant through all steps until a " +",

or " ++", or " +++" is printed in the PRC-column. (See the FEMO-part of the flowcharts in Section 2.4.1.) In single dropping, MS(2) is redefined at each step according to the redefinition of the model at each step in this ranking procedure.

"DF(2)". This is the number of degrees of freedom in the denominator of the value in the F-column. DF(2) is associated with MS(2) of the previous column in an obvious manner. If the two values are multiplied, the result is the residual sum of squares.

"COEFF DET". In this column the value of the coefficient of determination is printed for the model of a given step before the CIV or effect ranked at the step has been deleted. This means, for example, that the value of the coefficient of determination printed at the first step of the FCA for COMO is the one for the model containing all N IVs.

3.3.2 The Final FCA

In case there is more than one set of Control Cards 4 in a given problem when FEMO is used, the program will repeat, at the end of the printout of the problem, all FEMO FCAs which had been printed earlier at the ends of the printouts for each individual set of CC 4. The reason for printing the Final FCA is to ease the comparisons between the results corresponding to the various sets of CC 4.

In general, each FCA corresponding to a set of CC 4 will show a different significant model. It should be remembered that each set of CC 4 corresponds to a different model which includes one possible selection from the set of the confounded constants (see Appendix A). Consequently, there are, in general, as many models, or sets of CC 4 (and, therefore, as many FEMOs in the Final FCA), as there are possible selections from the set of the confounded constants. The Final FCA then serves in finding that significant model which contains the least number of significant effects which is then called "The most probable significant model." For further discussion of the use of the Final FCA the reader is referred to Example 5 in Section 3.4.5.

The sequence of the printout of the individual FCAs in the Final FCA is as follows, assuming the most general case where several dependent variables have been analyzed and where both dropping procedures have been performed (using the actual form of the printout in the identification of the individual FCAs):

```
FEMO/Y1/CUMUL/SET 1
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '

FEMO/Y1/CUMUL/SET 2
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '

FEMO/Y1/CUMUL/SET 3
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '

FEMO/Y1/SINGLE/SET 1
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '

FEMO/Y1/SINGLE/SET 2
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '

FEMO/Y1/SINGLE/SET 3
 "   /Y2/   "    /   "
 '    '     '        '
 '    '     '        '
```

Note. In the Final FCA, only the results of steps with accepted models are given. This is the only possible deviation from the printouts of the individual FCAs.

3.3.3 Additional Analyses of Variance ("ANVAs")

As mentioned earlier, in case of analysis of covariance when significant CIVs were found in COMO (cumulative dropping) and accordingly were kept in the model through the FEMO ranking, the program performs additional rankings of the factorial effects for the dependent variable(s) after exclusion of all CIVs from the model and for each significant OCIV and for each OCIV contained in a significant GCIV, provided the corresponding option (ADA > 0 in column 46 of CC No. 1) is requested. Only the cumulative ranking procedure is performed in the ANVAs which are actually FEMOs for y and the OCIVs concerned.

The ANVA for an OCIV may show factorial effects to be significant when the OCIV is not a "fixed" variate (i.e., does not correspond to the theory of analysis of covariance) but is a random variable itself like the dependent variable, y. For example, if one "factor" in the data layout is "time of the day" (with levels 8 a.m., 12 noon, 2 p.m., say), and if one covariate is air temperature, the factor "time of the day" will almost certainly appear to have an influence upon this OCIV.

If, in this example, the response variable y is a true function of the time of day (independent of air temperature) the factor "time of day" will appear to have a significant effect also upon y. Therefor, both the ANVAs for y and the OCIV (temperature) will show the effect "time of day" to be significant. If both variables are analyzed in combination in an analysis of covariance model, and if temperature exercises an additional effect upon the response variable, the OCIV "temperature" may be significant (in COMO); and if kept in the model through the FEMO ranking, may cause the effect "time of day" to be non-significant with respect to y. In other words, in performing an analysis of covariance alone there is the possible danger of not detecting the significance of a factorial effect. Performance of the ANVAs for y and the OCIVs will prevent the indicated danger. Also, the ANVAs will give, in combination with the analysis of covariance, a much better general picture of the relationship between the variables concerned than the analysis of covariance results alone could give. See also Example 6 in Section 3.4.6.

The final comprehensive printouts of the ANVAs are complemented by listings of the admissible effects at each step of an ANVA and the associated I(X)-values and their arguments. These complementary printouts serve, as they do in the other printouts, to inform the analyst how the ranking of the factorial effects was actually performed.

3.4 Examples of Application

In this section, 7 examples of application of the NOVACOM program are discussed. Since it is not possible to show all features of the program in the printout of one single example, those parts of the printout of the examples are reproduced which show features not exemplified in other examples. For Example 1, the complete printout is reproduced. For some examples, only the Final Comprehensive Analyses (FCAs) are reproduced.

Following is a list of the headings of the 7 examples:

Example 1: Multiple Regression (Duncan, 1959)

Example 2: Half Replicate of 2x2x2x2 Factorial (Davies, 1956)

Example 3: 3x4 Factorial (Hicks, 1964, and Robson, 1959)

A: Restricted Admissibility, Uncoded

B: Unrestricted Admissibility, Uncoded

C: Unrestricted Admissibility, Coded

Example 4: 2x4x4 Factorial, 5 cells empty (Stevens, 1948)

Example 5: 3x3x2 Factorial, 5 cells empty, 3 constants confounded

Example 6: 3x3x3 Factorial, 9 cells empty, with 3 CCIVs, 2 dependent variables

Example 7: 3x3x3 Factorial (Example 6 modified) with singularity in matrix A.

The areas of application exemplified are as follows:

I. Multiple (polynomial) regression: Example 1.

II. Analysis of variance for orthogonal data layouts: Examples 2, 3A, 3B, 3C.

III. Analysis of variance for non-orthogonal data layouts without confounding: Example 4.

IV. Analysis of variance for non-orthogonal data layouts with confounding: Example 5.

V. Analysis of covariance for non-orthogonal data layouts with confounding: Example 6.

The various features of the program are illustrated as follows:

1. Complete printout: Example 1.

2. Multiple dependent variables: Example 6.

3. Model generation options (not all possibilities illustrated):

CIVs, all automatically generated: Example 1.

CIVs, automatically generated; and deleted via CC No. 5: Example 6.

DIVs, all hand-generated via CC No. 4: Examples 6, 7.

DIVs, all automatically generated: Example 3.

DIVs, automatically generated; and hand-generated via CC No. 4: Examples 2, 4.

DIVs, automatically generated; and deleted via CC No. 4: Example 5.

4. Types of factors:

All qualitative: Examples 2, 4, 5.

All quantitative: Example 3.

Both qualitative and quantitative: Examples 6, 7.

5. Coding: Examples 3C, 6, 7.

6. Admissibility for ranking:

Restricted: Examples 1, 3A, 6, 7.

Unrestricted/Relaxed: Examples 3B, 3C.

7. ANVAs: Example 6.

All example problems were run with the option for both cumulative and single dropping and in all examples, except No. 1 and No. 2, 3 significance levels α were specified (1 only in Examples 1 and 2).

3.4.1 Example 1

This example is exhibited in order to show the capability of NOVACOM in multiple regression. The example also serves to illustrate the entire printout of the program for this case. The exhibit of the complete printout allows a comparison with the treatment of the problem by DA-MRCA (Reference 2). The same data was also used as an example in the documentation of DA-MRCA.

The data is taken from Duncan [1959], page 697. There are two OCIVs (x_1 = Plate Thickness in inches, and x_2 = Brinnell Hardness Number) and one dependent variable (y = Ballistic Limit in Feet/Sec.). A model of third order in x_1 and x_2 is automatically generated which leads to N=9 CIVs. The number of data points is n=20. See the reproduced input

sheet for this example. Following the input sheet is the reproduced printout from NOVACOM on which are written the numbers of the notes which follow below. The reader is also referred to the notes on the flowcharts in Section 2.4.2 which explain many of the features of the complete printout.

Notes on (complete) printout Example 1.

(References to "boxes" are to the comments in Section 2.4.2.)

Note 1.1. The entries on Control Card No. 1 are printed for identification purposes.

Note 1.2. The 2 OCIVs ("IV 1" and "IV 2") and the 7 automatically generated GCIVs (P=3, CC No. 1, columns 14-15) are identified. See Section 2.2.1 for the notation used.

Note 1.3. The data input is printed (x_1, x_2, y). See comments on Box 1.1.

Note 1.4. The maximum and minimum value for each OCIV is given plus the range and the interval size ("DELTA") for the frequency bar charts of the OCIV values. See comments on Box 1.1.

Note 1.5. The "FULL DATA MATRIX" contains the values of the 9 CIVs (2 OCIVs and 7 GCIVs) and of the dependent variable. The horizontal and vertical marginal identifications give the IV numbers and the observation numbers, respectively. See comments on Box 1.3.

Note 1.6. The "SUMMATION MATRIX" is the (N+2)x (N+2) - matrix composed of the (N+1) x (N+1) matrix (A) of the coefficients of the normal equations, i.e., of the terms $\sum_{i=1}^{n} x_{\nu i} x_{\nu' i}$, $\nu = 0,1,\ldots,N$, (N=9 here), with $x_{0i} = 1$; and of the N+2 terms (only the row vector printed) $\sum_{i=1}^{n} x_{\nu i} y_i$ and $\sum_{i=1}^{n} y_i^2$. The two marginal identifications give the IV numbers, $\nu = 1,\ldots,9$. See comments on Box 1.3.

Note 1.7. The averages of the values of the N=9 CIVs and of y_i are printed. (For example, $\bar{x}_1 = \frac{1}{20} \sum_{i=1}^{20} x_{1i} = 0.249799\ldots$, and $\bar{y} = \frac{1}{20} \sum_{i=1}^{20} y_i = 1179.15$.)

Note 1.8. This is a printout of the IV-numbers of the admissible CIVs at the first step of COMO. See Note 1.2: under restricted admissibility, only x_1^3, $x_1^2 x_2$, $x_1 x_2^2$, ar x_2^3 are admissible for ranking at the first step.

Note 1.9. The regression coefficients are always printed at the first step.

Note 1.10. The I(X)-value corresponding to the CIV (IV No. 6: x_1^3) ranked least important at the first step of COMO is printed together with its arguments: ARG 1 = $\frac{1}{2}f_2$=5 and ARG 2 = $\frac{1}{2}f_1$=0.50. (See also formula (2-12) in Section 2.1.2.)

Note 1.11. The printout identification indicates whether this is a COMO or a FEMO ranking; the number of the dependent variable is given ("Y1" if there is only one dependent variable as in the present example); the ranking option is indicated: CUMUL = cumulative dropping; SINGLE = single dropping; and the number of the SET of Control Cards No. 4 is given. If there is none or one Set of CC 4, or if the problem is one of multiple regression (as is the case here), "SET 1" is printed here. (See also the identification lines at previous places of the present example: "SET 1" is printed everywhere, whereas the other 3 spaces are left blank when not applicable.)

Note 1.12. The inverse matrix A^{-1} is printed for the first step of COMO where the model contains all N=9 IVs.

Note 1.13. The main diagonal elements of the computed identity matrix $A^{-1}A$ are not printed since all deviations from 1 were smaller than TOLI(2) = 0.001 which was used as input value.

Note 1.14. Following are the printouts associated with the Chi-square computations for the normality test of the residuals. These printouts are always given for the first (good) step.

Note 1.15. Admissible for ranking at the second step are CIVs Nos. 7, 8 and 9; that is, after dropping x_1^3 from the model, no CIV became additionally admissible for ranking. Following are the ranking informations for steps 2 through 7 (see Notes 1.8 and 1.10) and the values of the regression coefficients for each step because IRCO=1 in column 45 of CC No. 1.

Note 1.16. At Step Number 7, I(X) reaches for the first time, the first specified α-level: α_1 = 0.05, see columns 26-29 of CC No. 1. The "full printout" for this step follows; see the comments on Box 7.4.

Note 1.17. The statement "DEVIATIONS OF MAIN DIAGONAL IDENTITY MATRIX ELEMENTS LESS THAN .001" is printed, however, the actual computational check, in this example, was done for the first step only. Once the model of a step has been accepted by NOVACOM, the accuracy checks are not performed anymore after that step. See also Section 2.1.3.

Note 1.18. The "RESIDUAL OR ERROR SUM OF SQUARES", at this point of the "full printout" is defined as ATSS-ASSR(N'), with N'=9-(7-1)=3 in the present example. That is, should the analyst decide to use the model of this step (the "significant model" at α = 0.05, containing 3 CIVs) as the prediction model, while pooling all non-significant CIVs into the error, the value printed is the pooled error sum of squares. The "CHECK ERROR

SUM OF SQUARES" is the sum of the squares of the prediction errors given further below (see Note 1.19). It serves as an additional check on the computational accuracy. The "SQUARE ROOT OF (the) RESIDUAL VARIANCE" is the estimated error standard deviation, s, for this step based on the Error Sum of Squares discussed above. The value s is used in the computations of the standard deviations of the regression coefficients given later in the full printout (see Note 1.20). See also comments on Box 7.4.

Note 1.19. The "PREDICTED VALUES", the "PREDICTION ERRORS", and the subsequent Chi-square computations are based on the model of this step, that is, on the model containing the 3 significant CIVs. See Note 1.18.

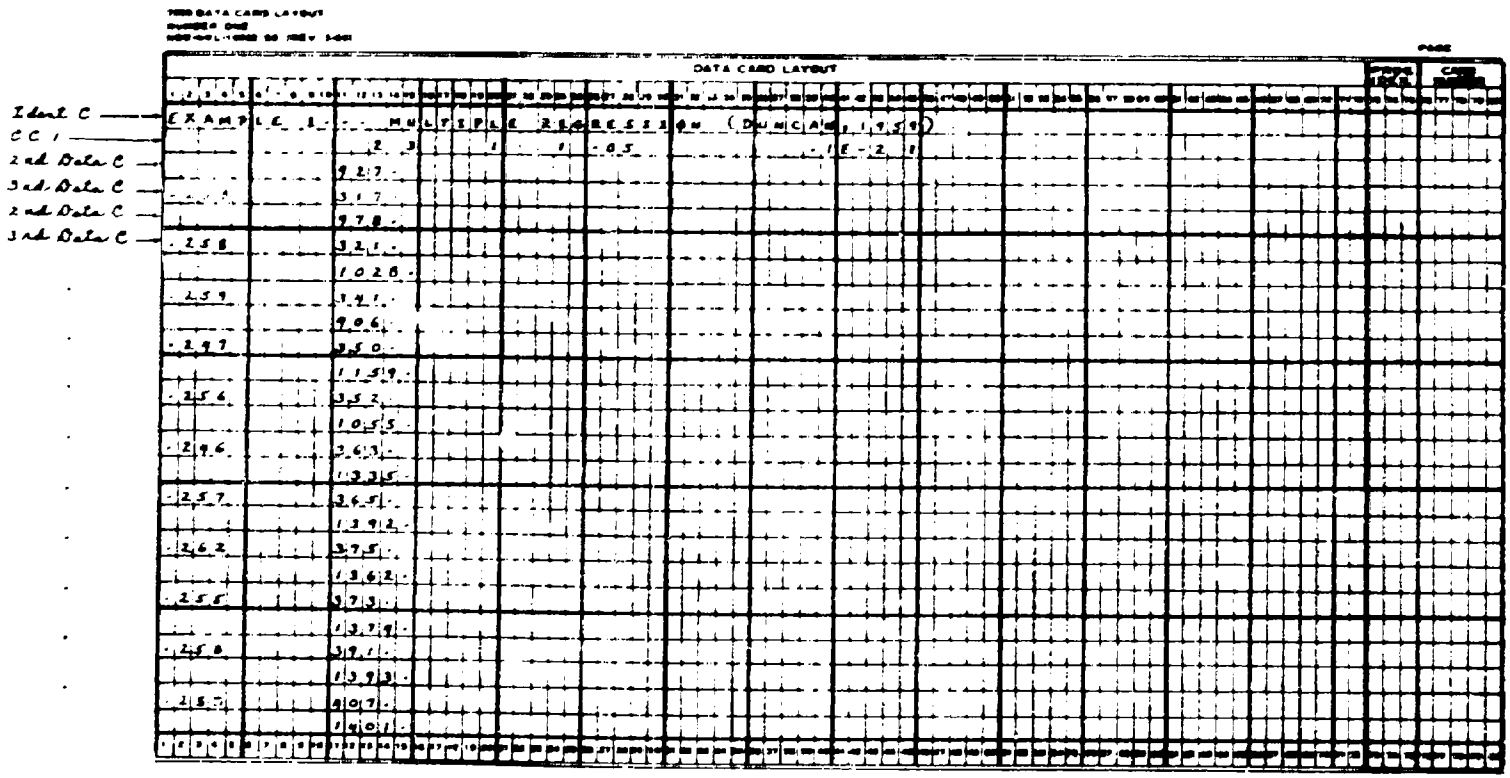
Note 1.20. The "STANDARD DEVIATIONS OF THE REGRESSION COEFFICIENTS" are the values $s\sqrt{c_{\nu\nu}}$, $\nu = 0,1,2,4$ in the present example, where s is the standard deviation of this step, see Note 1.18.

Note 1.21. The last 9 lines give the information on the I(X)-computations for COMO, single dropping.

Note 1.22. The FCA for COMO, cumulative dropping, shows that the significant model, at the $\alpha = 0.05$ level, based on this dropping procedure, contains x_1x_2, x_1, and x_2 (in their order of ranking, with x_2 being the most important CIV). See also Section 3.3.1 on the interpretation of the FCA results. A comparison with the treatment of the same problem by DA-MRCA (Reference 2) shows the same significant model obtained by NOVACOM and by the BIVOR option of DA-MRCA.

Note 1.23. The significant model, at the $\alpha = 0.05$ level, resulting from single dropping is the same as the one resulting from cumulative dropping.

Note 1.24. The indicated problem running time is 28 seconds. This is approximately the same time which the problem took when analyzed by DA-MRCA.


DATA CARD LAYOUT
Ident C
CC 1
2nd Data C
3rd Data C
2nd Data C
3rd Data C
EXAMPLE 1
MULTIPLE REGRESSION (DUMCAT, 1, 9, 9)

7000 DATA CARD LAYOUT
NUMBER ONE
NDW-NBL-10662-60 (REV. 3-64)

Input Example 1, continued

PAGE

DATA CARD LAYOUT

	1–5	11–15
	.252	926.
		1436.
	.246	432.
		1327.
	.250	969.
		950.
	.242	275.
		998.
	.243	302.
		1144.
	.239	331.
		1090.
	.242	355.
		1276.
	.244	385.
2nd Data C. →	LAST	1062.
3rd Data C. →	.234	926.

PROG IDEN. | CARD NUMBER

```
NOVACOM ... NON-ORTHOGONAL VARIANCE AND COVARIANCE ANALYSIS BY MULTIPLE REGRESSION ... PROGRAM VERSION OF JULY 15,1967

 F NO-ADV    NO  NCC4 NSCC4  TP  P GOC NCC5 CODE    C    DRAP  KALPHA ALPHA(1) ALPHA(2) ALPHA(3)  CAP    TOLT2    IRCV ADA
 0    1    0        0    1    2    0      0    1   1.00    1      1    .0500    .0000    .0000    0   .1000E-02    1    0
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/    /    /      /SET  1

IDENTIFICATION OF INDEPENDENT VARIABLES (IV-S)

CIV-S ..
        IV   1      1( 1)
        IV   2      2( 1)
        IV   3      1( 2)
        IV   4      1( 1)X 2( 1)
        IV   5      2( 2)
        IV   6      1( 3)
        IV   7      1( 2)X 2( 1)
        IV   8      1( 1)X 2( 2)
        IV   9      2( 3)
```

Note 1.1

Note 1.2

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/    /    /      /SET  1

     DATA INPUT
                 (  1)                 (  2)
                 ( Y1)

   1             0.253000000          317.000000000
                927.000000000

   2             0.258000000          321.000000000
                978.000000000

   3             0.259000000          341.000000000
               1028.000000000

   4             0.247000000          350.000000000
                906.000000000

   5             0.256000000          352.000000000
               1159.000000000

   6             0.246000000          363.000000000
               1035.000000000

   7             0.257000000          365.000000000
               1335.000000000

   8             0.262000000          375.000000000
               1392.000000000
```

Note 1.3

Data Input, (Example 1), continued:

9	0.255000000 1362.000000000	373.000000000
10	0.258000000 1374.000000000	391.000000000
11	0.253000000 1393.000000000	407.000000000
12	0.252000000 1401.000000000	426.000000000
13	0.246000000 1436.000000000	432.000000000
14	0.258000000 1327.000000000	469.000000000
15	0.242000000 950.000000000	275.000000000
16	0.243000000 998.000000000	302.000000000
17	0.239000000 1144.000000000	331.000000000
18	0.242000000 1080.000000000	355.000000000
19	0.244000000 1276.000000000	385.000000000
20	0.234000000 1062.000000000	426.000000000
	(1)	(2)
MAX	0.262000000	469.000000000
MIN	0.234000000	275.000000000
RANGE	0.028000000	194.000000000
DELTA	0.000933333	6.466666667

Note 1.4

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/    / /      /SET  1

FREQUENCY BAR CHART FOR RECIV NO.  1

     UPPER BOUND  FREQUENCY     BAR CHART
          0.2349          1    IX
          0.2359          0    I
          0.2368          0    I
          0.2377          0    I
          0.2387          0    I
          0.2396          1    IX
          0.2405          0    I
          0.2415          0    I
          0.2424          2    IXX
          0.2433          1    IX
          0.2443          1    IX
          0.2452          0    I
          0.2461          2    IXX
          0.2471          1    IX
          0.2480          0    I
          0.2489          0    I
          0.2499          0    I
          0.2508          1    IX
          0.2517          0    I
          0.2527          1    IX
          0.2536          2    IXX
          0.2545          0    I
          0.2555          1    IX
          0.2564          1    IX
          0.2573          1    IX
          0.2583          2    IXX
          0.2592          1    IX
          0.2601          0    I
          0.2611          0    I
          0.2620          1    IX

X REPRESENTS    1.0000 OBSERVATIONS.
```

```
EXAMPLE 1.,. MULTIPLE REGRESSION (DUNCAN,1959)
/    / /      /SET  1

FREQUENCY BAR CHART FOR ACTV NO.  2

     UPPER BOUND  FREQUENCY     BAR CHART
        281.4667          1    IX
        287.9333          0    I
        294.4000          0    I
        300.8667          0    I
        307.3333          1    IX
        313.8000          0    I
        320.2667          1    IX
        326.7333          1    IX
        333.2000          1    IX
        339.6667          0    I
        346.1333          1    IX
        352.6000          2    IXX
        359.0667          1    IX
        365.5333          2    IXX
        372.0000          0    I
        378.4667          2    IXX
        384.9333          0    I
        391.4000          2    IXX
        397.8667          0    I
        404.3333          0    I
        410.8000          1    IX
        417.2667          0    I
        423.7333          0    I
        430.2000          2    IXX
        436.6667          1    IX
        443.1333          0    I
        449.6000          0    I
        456.0667          0    I
        462.5333          0    I
        469.0000          1    IX

X REPRESENTS    1.0000 OBSERVATIONS.
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1950)
/    /  /      /SET  1
```

Note 1.5

	FULL DATA MATRIX (1) (7) (Y1)	(2) (8)	(3) (9)	(4)	(5)	(6)
1	.253800E+00 .252900E+02 927.000000000	.317000E+03 .254257E+05	.640090E-01 .318550E+00	.802010E+02	.100489E+06	.161943E-01
2	.258800E+00 .213670E+02 978.000000000	.321000E+03 .265846E+05	.665644E-01 .330762E+00	.828180E+02	.103041E+06	.171735E-01
3	.259000E+00 .228746E+02 1028.000000000	.341000E+03 .301168E+05	.670816E-01 .396518E+00	.883190E+02	.116281E+06	.173740E-01
4	.247800E+00 .213531E+02 906.000000000	.350000E+03 .302575E+05	.610099E-01 .428750E+00	.866500E+02	.122900E+06	.150692E-01
5	.256000E+00 .230687E+02 1150.000000000	.352000E+03 .317194E+05	.655936E-01 .430142E+00	.901120E+02	.123904E+06	.167772E-01
6	.246000E+00 .219673E+02 1059.000000000	.383000E+03 .321152E+05	.605160E-01 .478321E+00	.892980E+02	.131769E+06	.148809E-01
7	.257000E+00 .241879E+02 1339.000000000	.365000E+03 .342386E+05	.660409E-01 .486271E+00	.938090E+02	.133229E+06	.169746E-01
8	.262000E+00 .257415E+02 1392.000000000	.375000E+03 .368638E+05	.688440E-01 .527344E+00	.982300E+02	.140625E+06	.179847E-01
9	.255000E+00 .242543E+02 1362.000000000	.373000E+03 .354778E+05	.650250E-01 .518895E+00	.951190E+02	.139129E+06	.165814E-01
10	.258800E+00 .260895E+02 1374.000000000	.391000E+03 .394433E+05	.665644E-01 .567769E+00	.1008780E+03	.1920815E+06	.171736E-01
11	.253000E+00 .260517E+02 1393.000000000	.407000E+03 .419092E+05	.640090E-01 .674194E+00	.1029118E+03	.169644E+06	.161943E-01
12	.252000E+00 .270927E+02 1401.000000000	.426000E+03 .497320E+05	.635040E-01 .773084E+00	.1073932E+03	.181476E+06	.160000E-01
13	.246000E+00 .261896E+02 1436.000000000	.432000E+03 .490095E+05	.605160E-01 .806814E+00	.1062728E+03	.186634E+06	.148809E-01

Full Data Nature. Example 1. Continued

14	.250000E+00 .293125E+02 1327.000000000	.469000E+03 .549903E+05	.625000E-01 .193162E+09	.117250E+03	.220961E+08	.156250E+01
15	.242000E+00 .161851E+02 959.000000000	.275000E+03 .183012E+05	.585640E-01 .207960E+08	.665500E+02	.756250E+06	.141729E+01
16	.243000E+00 .178528E+02 998.000000000	.302000E+03 .221626E+05	.590490E-01 .275434E+08	.733660E+02	.912040E+06	.143489E+01
17	.239000E+00 .189871E+02 1144.000000000	.331000E+03 .261851E+05	.571210E-01 .362647E+08	.791090E+02	.106561E+08	.136519E+01
18	.242000E+00 .207902E+02 1080.000000000	.355000E+03 .304980E+05	.585640E-01 .447380E+08	.858100E+02	.126025E+08	.141729E+01
19	.244000E+00 .229214E+02 1276.000000000	.385000E+03 .361649E+05	.595360E-01 .570664E+08	.939480E+02	.148229E+08	.145268E+01
20	.234000E+00 .233261E+02 1062.000000000	.426000E+03 .[illegible]E+05	.5[illegible]E-01 .773088E+08	.996840E+02	.181476E+08	.128120E+01

EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/ / / /SET 1

SUMMATION MATRIX (NORMAL EQUATIONS)

Note 1.6

	(0)	(1)	(2)	(3)	(4)	(5)
	(6)	(7)	(8)	(9)		
(0)	0.20000000000000E+02	0.49999999999999E+01	0.73560000000000E+04	0.12491160000000E+01	0.18376700000000E+04	0.27496700000000E+07
	0.31258406199999E-00	0.45949424599999E+03	0.68884107799998E+06	0.10441680840000E+10		
(1)	0.49999999999999E+01	0.12491160000000E+01	0.18376700000000E+04	0.31258406199999E-00	0.45949424599999E+03	0.68884107799998E+06
	0.78290694647998E-01	0.11499383634800E+03	0.17171842192999E+06	0.26075072878999E+09		
(2)	0.73560000000000E+04	0.18376700000000E+04	0.27496700000000E+07	0.45949424599999E+03	0.68884107799998E+06	0.10441680840000E+10
	0.11499383634800E+03	0.17171842192999E+06	0.26075072878999E+09	0.40262394985400E+12		
(3)	0.12491160000000E+01	0.31258406199999E-00	0.45949424599999E+03	0.78290694647997E-01	0.11499383634800E+03	0.17171842193000E+06
	0.19625828659605E-01	0.28803545459713E+02	0.42969185981099E+05	0.65172278713992E+08		
(4)	0.18376700000000E+04	0.45949424599999E+03	0.68884107799998E+06	0.11499383634800E+03	0.17171842192999E+06	0.26075072878999E+09
	0.28803545459713E+02	0.42969185981099E+05	0.65172278713992E+08	0.10050289977022E+12		
(5)	0.27496700000000E+07	0.68884107799998E+06	0.10441680840000E+10	0.17171842193000E+06	0.26075072878999E+09	0.40262394985400E+12
	0.42969185981099E+05	0.65172278713992E+08	0.10050289977022E+12	0.15755163620368E+15		
(6)	0.31258406199999E-00	0.78290694647998E-01	0.11499383634800E+03	0.19625828659605E-01	0.28803545459713E+02	0.42969185981099E+05
	0.49239705303982E-02	0.72208612249340E+01	0.10761466839659E+05	0.16303347525265E+08		
(7)	0.45949424599999E+03	0.11499383634800E+03	0.17171842192999E+06	0.28803545459713E+02	0.42969185981099E+05	0.65172278713992E+08
	0.72208612249340E+01	0.10761466839659E+05	0.16303347525264E+08	0.25109381702289E+11		
(8)	0.68884107799998E+06	0.17171842192999E+06	0.26075072878999E+09	0.42969185981099E+05	0.65172278713992E+08	0.10050289977022E+12
	0.10761466839659E+05	0.16303347525264E+08	0.25109381702289E+11	0.39308453181165E+14		
(9)	0.10441680840000E+10	0.26075072878999E+09	0.40262394985400E+12	0.65172278713992E+08	0.10050289977022E+12	0.15755163620368E+15
	0.16303347525265E+08	0.25109381702289E+11	0.39308453181165E+14	0.62526526242079E+17		
(Y1)	0.23583000000000E+05	0.59002529999999E+04	0.87957870000000E+07	0.14774739810000E+04	0.22007870068999E+07	0.33312778490000E+10
	0.37029302081699E+03	0.95112250611498E+06	0.83340633016498E+09	0.12804960670510E+13	0.23668483000000E+08	

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/    / /      /SET  1

      AVERAGES                                                                  Note 1.7

(  1)  0.24979999999999E+00 (  2)  .36780000000000E+03 (  3)  0.62455799999998E-01 (  4)  0.91883490999997E+02
(  5)  0.13748350000000E+06 (  6)  0.15629203100000E+01 (  7)  0.22974712299999E+02 (  8)  0.34342053899999E+05
(  9)  0.52208404200000E+08

   Y1  0.11791500000000E+04

      TOTAL SUM OF SQUARES ADJUSTED FOR THE MEAN

   Y1  0.66058955000019E+06
THE FOLLOWING CIV.S ARE ADMISSIBLE,.,/ 6/ 7/ 8/ 9/                                Note 1.8

REGRESSION  COEFFICIENTS.STEP   1 /COMB/Y1/CUMUL /SET  1                          Note 1.9
(  0)  0.12821315885690E+07 (  1)  -.11615967250284E+08 (  2)  -.24858970116104E+04 (  3)  0.33453352065839E+08
(  4)  0.17033208982302E+05 (  5)  0.97098933633693E+00 (  6)  -.29537023939548E+08 (  7)  -.29518659455084E+05
(  8)  -.29323499306479E+01 (  9)  -.23700792204804E+03
ARG1=   5.00  ARG2=   0.50  X= .97368559E+00  I(X)= 0.61447263797231E+00                Note 1.10
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET  1   Note 1.11

STEP NUMBER   1   DETERMINANT=      314.21608595135   INVERSION TIME=   0.0481 SECONDS

INVERSE MATRIX                Note 1.12
                 ( 0)                ( 1)                ( 2)                ( 3)                ( 4)                ( 5)
                 ( 6)                ( 7)                ( 8)                ( 9)

( 0) 0.16300842611634E+09 -.17808285017171E+10 -.12353751988268E+06 0.66740700611147E+10 0.90844456813198E+06 0.28682492654521E+02
     -.78490587069814E+10 -.16363379503610E+07 -.12491560475398E+03 0.19468264283319E+02

( 1) -.17808280583348E+10 0.19723700541956E+11 0.11640847259039E+07 -.72713553737218E+11 -.86815990489631E+07 -.22443965229281E+03
     0.89357028649998E+11 0.19940939832463E+08 0.96974628651053E+03 -.93435014471857E+02

( 2) -.12353753384482E+06 0.11640884918175E+07 0.22326959039667E+03 -.39437946447622E+07 -.15457988995678E+04 -.87810363303220E+01
     0.34964103131883E+07 0.29415659662403E+04 0.39049459744335E+00 -.73139220394977E+05

( 3) 0.64740669189316E+10 -.72713537036466E+11 -.35437797799971E+07 0.27193126245742E+12 0.26796064467385E+08 0.53829990619269E+03
     -.33894755945892E+12 -.50063464046977E+08 -.23267904116471E+04 0.18196198261585E+01

( 4) 0.90844649587483E+06 -.96016225589433E+07 -.15457986263986E+04 0.26796158462922E+08 0.10928632900920E+05 0.53066431329498E+00
     -.26637370841414E+08 -.18621381982422E+05 -.23169072236086E+01 0.34142731576122E+04

( 5) 0.28682701902212E+02 -.22444202426996E+03 -.87810461422886E+01 0.53830483418485E+03 0.53066509787688E+00 0.61090226969009E+04
     -.43374575103578E+03 -.69068036713231E+00 -.28539799852856E+03 0.85516654943976E+08

( 6) -.78495334040086E+10 0.88356991428376E+11 0.34963910860554E+07 -.33894749801019E+12 -.26337268142620E+08 -.43373460548146E+03
     0.42879839249482E+12 0.58022504169118E+08 0.20148669476942E+04 -.40672760310084E+01

( 7) -.16363434388417E+07 0.15940971404518E+08 0.29415847450339E+04 -.50063994043708E+08 -.18621376633588E+05 -.69067878038610E+00
     0.50022677383732E+08 0.33748210120461E+05 0.25812175742486E+01 0.17250618039891E+04

( 8) -.12491656290321E+03 0.94975710870615E+03 0.39049507620386E+00 -.23268310075505E+04 -.23169109956781E+01 -.28539801332115E+03
     0.20149174751328E+04 0.25812251033943E+01 0.14183919382029E+02 -.57659315322613E+07

( 9) 0.19468507823586E+02 -.93437678933316E+02 -.73139402223355E+05 0.18197167360809E+01 0.34142866759263E+04 0.83516692276197E+06
     -.40673933418912E+01 0.17250369707925E+04 -.57659330706612E+07 0.52655683343253E+11

DEVIATIONS OF MAIN DIAGONAL IDENTITY MATRIX ELEMENTS LESS THAN .00100000 .   Note 1.13
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET  1

STEP NUMBER   1

REGRESSION SUM OF SQUARES ADJUSTED FOR THE MEAN =        585304.00996113

RESIDUAL OR ERROR SUM OF SQUARES =        75284.540039063

CHECK ERROR SUM OF SQUARES =        75284.635627900

SQUARE ROOT OF RESIDUAL VARIANCE=       84.766664128029
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET  1

STEP NUMBER   1
```

Note 1.14

ITEM NUMBER	PREDICTED VALUE	PREDICTION ERROR			
1.	848.974423846	0.784255761559478E+02	2.	967.463079100	0.1[illegible]E+02
3.	1145.530225159	-.117530229159968E+03	4.	1049.681891250	-.143[illegible]E+03
5.	1165.374814383	-.637481438327818E+01	6.	1120.319368575	-.653[illegible]E+02
7.	1266.575971052	0.684240289479498E+02	8.	1[illegible].641508719	0.2[illegible]E+02
9.	1285.176052817	0.768239470827398E+02	10.	1302.466124106	-.184[illegible]E+02
11.	1412.098138042	-.190981380417368E+02	12.	1430.965449309	-.379[illegible]E+02
13.	1379.042021460	0.569579785396348E+02	14.	1336.547909033	-.954[illegible]E+01
15.	997.259567887	-.472595672865578E+02	16.	931.006830055	0.7[illegible]E+02
17.	1115.381114076	0.286188859240508E+02	18.	1088.084380709	-.808[illegible]E+01
19.	1210.452655053	0.655473440468782E+02	20.	1076.362093940	-.143[illegible]E+02

EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1950)
/CONT/V1/CUMUL /SET 1

STEP NUMBER 1

PREDICTION ERROR FREQUENCY DISTRIBUTION FOR THE 20 OBSERVATIONS

RANGE= 0.22210746741400E+03

UPPER BOUND	FREQUENCY	BAR CHART	CHI SQU CONTR	OBS PR	THEO PR
-136.2793	1.	I*			
-129.8747	0.	I			
-121.4711	0.	I			
-114.0674	1.	I*			
-106.6640	0.	I			
-99.2604	0.	I			
-91.8568	0.	I			
-84.4532	0.	I			
-77.0497	0.	I			
-69.6461	0.	I			
-62.2425	1.	I*			
-54.8389	0.	I	0.980	5.0	5.274
-47.4353	0.	I			
-40.0317	1.	I*			
-32.6282	1.	I*			
-25.2246	0.	I			
-17.8210	2.	I**			
-10.4174	1.	I*			
-3.0138	3.	I***			
4.3898	3.	I	1.804	8.0	5.139
11.7933	1.	I*			
19.1969	0.	I			
26.6005	1.	I*			
34.0041	1.	I*			
41.4077	0.	I			
48.8112	0.	I			
56.2148	0.	I			
63.6184	1.	I*			
71.0220	2.	I**			
78.4256	3.	I***	0.037	9.0	9.997

* REPRESENTS 1.0000 OBSERVATIONS.

CHISQUARE COULD NOT BE COMPUTED.

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET  1

STEP NUMBER   1

STANDARD DEVIATION OF REGRESSION COEFFICIENTS

(  0)        1107791.2430896   (  1)        12185605.127237   (  2)        1296.4859732990   (  3)        49246237.430695
(  4)        9370.5959076649   (  5)        0.67794824703971  (  6)        56817134.002996   (  7)        13939.691951008
(  8)        3.2677665868712   (  9)        0.00019910195632
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 7/ 8/ 9/   Note 1.15

REGRESSION  COEFFICIENTS,STEP   2 /COMB/Y1/CUMUL /SET  1
(  0)  0.7414695907632E+06 (  1)  -.5460761010802?E+07 (  2)  -.2245052?612966E+04 (  3)  0.1010552855839?E+08
(  4)  0.1519833865055?2E+05 (  5)  0.941111501392?6E+00 (  7)  -.2607293?794298E+05 (  8)  -.2793555?831438E+01
(  9)  -.2358096?647446E+03
ARG1=   5.00  ARG2=   1.00  X= .9?859610?+00  1(X)= 0.619233?3490?490E+00
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 7/ 9/

REGRESSION  COEFFICIENTS,STEP   3 /COMB/Y1/CUMUL /SET  1
(  0)  0.5669190762?2?2E+06 (  1)  -.4?705?5?50136?2E+07 (  2)  -.150337190006?0E+04 (  3)  0.864994706?67473E+07
(  4)  0.1085?64493465?E+05 (  5)  0.?7927?7?776644E+00 (  7)  -.2142105?100095E+05 (  8)  -.353755?3819184E+01
ARG1=   5.00  ARG2=   1.90  X= .?3828?36E+00  1(X)= 0.3652495?524828E+00
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 3/ 4/ 9/

REGRESSION  COEFFICIENTS,STEP   4 /COMB/Y1/CUMUL /SET  1
(  0)  0.5464250?13668E+09 (  1)  -.2647948538586BE+06 (  2)  -.1552584609838?E+03 (  3)  0.351095886?0?0E+06
(  4)  0.2658?7?04208?1?E+03 (  5)  0.2?68?962?17?00E+00 (  9)  -.2?3?34860?0350E+03
ARG1=   5.00  ARG2=   2.00  X= .69860614E+00  1(X)= 0.417598858?42?8E-00
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 4/ 9/

REGRESSION  COEFFICIENTS,STEP   5 /COMB/Y1/CUMUL /SET  1
(  0)  0.3125102932?180E+05 (  1)  -.?7?9?020015?46E+05 (  2)  -.1582688?802524E+03 (  4)  0.29287?80346?58E+03
(  5)  0.2950?122829840E+00 (  9)  -.2?69961937?631E+03
ARG1=   5.00  ARG2=   2.?? X= .52862?91E+00  1(X)= 0.20?95895968467E-00
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 4/ 9/

REGRESSION  COEFFICIENTS,STEP   6 /COMB/Y1/CUMUL /SET  1
(  0)  0.2033820568??41E+05 (  1)  -.8956647429?403E+05 (  2)  -.5119295?0?63268E+02 (  4)  0.2610049?69?8?85E+03
(  5)  -.1410052?5440886E+01
ARG1=   5.00  ARG2=   3.00  X= .44993293E-00  1(X)= 0.1528?982234567E-00
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 4/

REGRESSION  COEFFICIENTS,STEP   7 /COMB/Y1/CUMUL /SET  1
(  0)  0.1834515837140?E+05 (  1)  -.?372?2387960?6E+05 (  2)  -.5215?67?795802E+02 (  4)  0.2223829439?7198E+03
ARG1=   5.00  ARG2=   3.50  X= .296852302E-00  1(X)= 0.4008?011824592E-01
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/V1/CUMUL /SET  1
                                                  Note 1.16
STEP NUMBER    7    DETERMINANT=      1717.0871573728    INVERSION TIME=    0.0073 SECONDS

INVERSE MATRIX
                 (  0)                  (  1)                  (  2)                  (  4)

(  0) 0.4763974696337E+04 -.1928686319774 3E+05 -.1282538479387 2E+02 0.5192644819732 3E+02

(  1) -.1928686319774 3E+05 0.7813364712757 4E+05 0.5192875447261 3E+02 -.2103785527306 4E+03

(  2) -.1282559475393 0E+02 0.5192875447304 6E+02 0.3493997650189 0E-01 -.1414523986442 0E+00

(  4) 0.5192644819755 5E+02 -.2103785527315 7E+03 -.1414523986442 0E-00 0.5970333130828 7E+00

DEVIATIONS OF MAIN DIAGONAL IDENTITY MATRIX ELEMENTS LESS THAN .00100000 .  Note 1.17
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/V1/CUMUL /SET  1

STEP NUMBER    7

REGRESSION SUM OF SQUARES ADJUSTED FOR THE MEAN =        493264.47243023

RESIDUAL OR ERROR SUM OF SQUARES =        167324.07756996

CHECK ERROR SUM OF SQUARES =        167324.07761057          Note 1.18

SQUARE ROOT OF RESIDUAL VARIANCE=        102.26314466902
```

```
EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/V1/CUMUL /SET  1

STEP NUMBER    7

ITEM NUMBER      PREDICTED VALUE        PREDICTION ERROR        Note 1.19

      1.          994.811713100      -.4781171309994 5E+02      2.          999.544932272      -.2134453227326 6E+02
      3.         1106.014112269      -.7801411226857 4E+02      4.         1105.648165403      -.1996481654029 3E+03
      5.         1152.191781418      0.4808218581951 2E+01      6.         1134.679787243      -.7967978724348 9E+02
      7.         1221.690513310      0.1133094866897 4E+03      8.         1319.998627261      0.7200137273850 6E+02
      9.         1243.204345120      0.1187956548801 2E+03     10.         1364.805354432      0.9194645967707 1E+01
     11.         1364.360445382      0.2863955441792 6E+02     12.         1421.364303534      -.2036430353402 2E+02
     13.         1310.589520062      0.1254104799380 3E+03     14.         1527.215375195      -.2002153751045 6E+03
     15.          960.602859321      -.1083285932114 7E+02     16.          998.856154930      -.8561549295594 6E+00
     17.         1053.899852302      0.9010014769772 5E+02     18.         1093.393630667      -.1339363064747 5E+02
     19.         1166.978559177      0.1090214408228 6E+03     20.         1043.150277680      0.1884972248030 6E+02
```

EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMM/Y1/CUMUL /SET 1

STEP NUMBER 7

PREDICTION ERROR FREQUENCY DISTRIBUTION FOR THE 20 OBSERVATIONS

RANGE= 0.325625895132594E+03

UPPER BOUND	FREQUENCY	BAR CHART	CHI SQU CONTR	OBS FR	THEO FR
-189.3612	2.	IXX			
-178.5070	0.	I			
-167.6528	0.	I			
-156.7986	0.	I			
-145.9444	0.	I			
-135.0902	0.	I			
-124.2360	0.	I			
-113.3818	0.	I			
-102.5276	0.	I			
-91.6734	0.	I			
-80.8192	0.	I			
-69.9650	2.	IXX			
-59.1108	1.	IX	0,071	5,0	5,632
-48.2566	0.	I			
-37.4024	0.	I			
-26.5483	0.	I			
-15.6941	2.	IXX			
-4.8399	2.	IXX			
6.0143	1.	IX			
16.8685	2.	IXX	0,308	7,0	5,678
27.7227	1.	IX			
38.5769	1.	IX			
49.4311	0.	I			
60.2853	0.	I			
71.1395	0.	I			
81.9937	1.	IX			
92.8479	1.	IX			
103.7021	0.	I			
114.5563	2.	IXX			
125.4105	2.	IXX	0,095	8,0	8,690

X REPRESENTS 1.0000 OBSERVATIONS.

CHISQUARE COULD NOT BE COMPUTED.

```
EXAMPLE 1.,. MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET  1

STEP NUMBER   7

STANDARD DEVIATION OF REGRESSION COEFFICIENTS
```

Note 1.20

```
(  0)        7058.3623792559  (  1)        28585.005283363  (  2)        19.115274223927  (  4)        77.412156128055
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 1/ 2/

REGRESSION  COEFFICIENTS.STEP   8 /COMB/Y1/CUMUL /SET  1
(  0)  -.18064735496999E+04 (  1)  0.79205343812347E+04 (  2)  0.27381023933320E+01
ARG1=   5.00  ARG2=   4.00  X= .23267561E-00  I(X)= 0.20104708809802E-01
THE FOLLOWING CIV-S ARE ADMISSIBLE,../ 2/

REGRESSION  COEFFICIENTS.STEP   9 /COMB/Y1/CUMUL /SET  1
(  0)  0.16275551965432E+03 (  2)  0.27834433940883E+01
ARG1=   5.00  ARG2=   4.50  X= .11396566E-00  I(X)= 0.11603335716437E-02
ARG1=   5.00  ARG2=   0.50  X= .97368559E+00  I(X)= 0.61447263797231E+00
ARG1=   5.50  ARG2=   0.50  X= .93315142E+00  I(X)= 0.39369799719664E-00
ARG1=   6.00  ARG2=   0.50  X= .81255176E+00  I(X)= 0.12202392750570E-00
ARG1=   6.50  ARG2=   0.50  X= .94653093E+00  I(X)= 0.40699267858731E-00
ARG1=   7.00  ARG2=   0.50  X= .75646967E+00  I(X)= 0.52068540479009E-01
ARG1=   7.50  ARG2=   0.50  X= .85113662E+00  I(X)= 0.12612087329659E-00
ARG1=   8.00  ARG2=   0.50  X= .65972626E+00  I(X)= 0.11048871698524E-01
ARG1=   8.50  ARG2=   0.50  X= .78386215E+00  I(X)= 0.44899259534230E-01
ARG1=   9.00  ARG2=   0.50  X= .48980578E-00  I(X)= 0.40332896718453E-03
```

Note 1.21

EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/CUMUL /SET 1

FINAL COMPREHENSIVE COMB-CUMULATIVE DROPPING

Note 1.22

STEP	CONC VAR	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
1	1(3)		0.61447264	2034.6074	1	.27025568	2034.6074	1	7529.4540	10	.88033414
2	1(1)X 2(2)		0.61923335	5538.9456	1	.50299924	3780.7745	2	7528.4540	10	.88295415
3	1(2)X 2(1)		0.86524958	19114.402	1	1.1816572	8896.0518	3	7528.4540	10	.87456929
4	1(2)		0.41759886	5760.3882	1	1.0775301	8112.1359	4	7528.4540	10	.84563357
5	2(3)		0.20395896	34682.557	1	1.7833967	13426.230	5	7528.4540	10	.83491349
6	2(2)		0.1528398Z	24908.425	1	2.0375924	15339.921	6	7528.4540	10	.78441098
7	1(1)X 2(1)	*	0.04008601	86302.441	1	3.3841508	25477.424	7	7528.4540	10	.74670456
8	1(1)		0.02010471	69933.580	1	4.1222864	31034.443	8	7528.4540	10	.81609979
9	2(1)		0.00116033	337028.45	1	8.6583974	65033.778	9	7528.4540	10	.91019422

EXAMPLE 1... MULTIPLE REGRESSION (DUNCAN,1959)
/COMB/Y1/SINGLE/SET 1

FINAL COMPREHENSIVE COMB-SINGLE DROPPING

Note 1.23

STEP	CONC VAR	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
1	1(3)		0.61447264	2034.6074	1	.27025568	2034.6074	1	7528.4540	10	.88033414
2	1(1)X 2(2)		0.39369400	5538.9456	1	.78801181	5538.9456	1	7029.0134	11	.88295415
3	1(2)X 2(1)		0.12202353	19114.402	1	2.7682900	19114.402	1	6904.8411	12	.87456929
4	1(2)		0.40699268	5760.3882	1	.73438371	5760.3882	1	7844.0535	13	.84563357
5	2(3)		0.05204854	34682.557	1	4.9070258	34682.557	1	7695.2203	14	.83491349
6	2(2)		0.12612087	24908.425	1	2.6234926	24908.425	1	8494.3760	15	.78441098
7	1(1)X 2(1)	*	0.01104887	86302.441	1	8.2524833	86302.441	1	10497.754	16	.74670456
8	1(1)		0.04489926	69933.580	1	4.6874867	69933.580	1	14919.206	17	.81609979
9	2(1)		0.00040333	337028.45	1	18.749260	337028.45	1	17975.960	18	.91019422

Note 1.24

PROBLEM RUNNING TIME(HRS.MIN.SEC)=00:00:28

3.4.2 Example 2

Example 2 is to show the capability of NOVACOM in the analysis of variance for incomplete but balanced data classifications. This capability is demonstrated with one of the simplest cases possible, namely that of a half-replicate of a 2^4-factorial experiment. The data is taken from Davies [1956] p. 455. The layout of the 8 observations is given in Table 3.1. (The cell numbers are indicated in the upper left corners of the cells.) On the reproduced input sheet for this example note that only 2 of the 3 two-factor interactions have been fitted in order to provide the minimum of 1 degree of freedom for error in the FEMO ranking. The two interaction effects fitted are $\mathcal{AB}$ and $\mathcal{AC}$; their respective aliases $\mathcal{CB}$ and $\mathcal{BB}$ could have been fitted as well. The four main effects are automatically generated whereas the two interactions are written on two CCs No. 4, that is, they have been "hand"-generated.

		C_1		C_2	
		B_1	B_2	B_1	B_2
A_1	B_1	1111 107			1122 108
	B_2		1212 122	1221 120	
A_2	B_1		2112 114	2121 121	
	B_2	2211 130			2222 132

Table 3.1

Data Layout Example 2

A partial reproduction of the printout of this example is given in order to show features which could not be shown with Example 1. The notes on these features follow below.

Notes on printout Example 2.

Note 2.1. The numbers printed here are programming information on the admissibility of effects; see Herring [1967].

Note 2.2. The DIVs and the effects are identified. In the present case the number of DIVs equals the number of the effects since all factorial effects are one-degree-of-freedom effects.

Note 2.3. For a crossed data classification the FULL DATA MATRIX also contains the cell identifications. The numerical values of IVs No. 1 through 6 are the design point coordinates which values are either 1 or 0.

Note 2.4. At the first step (of FEMO), effects No. 4, 5, and 6 are admissible for ranking, that is, $\mathcal{B}$, $\mathcal{AB}$, and $\mathcal{AC}$. $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are sub-effects of $\mathcal{AB}$ and $\mathcal{AC}$ and, therefore, are not admissible at the first step.

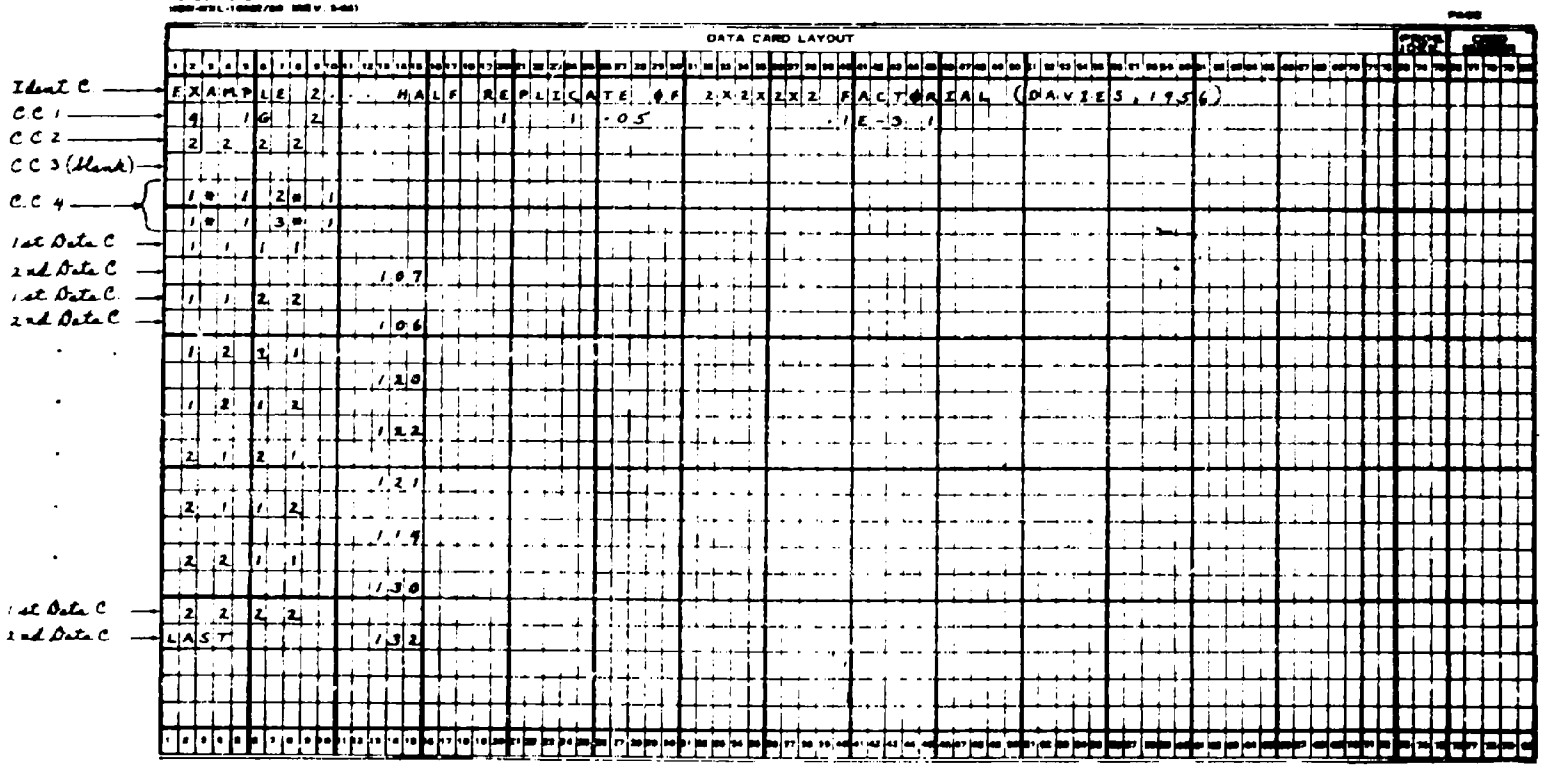
Note 2.5. The three I(X)-values corresponding to the three effects 4, 5, and 6 are printed. The second I(X)-value, that is, the I(X)-value corresponding to effect No. 5, is the largest one. Therefore, effect No. 5 ($\mathcal{AB}$) is ranked as the least important and deleted from the model. With $\mathcal{AB}$ deleted from the model, effect No. 2 ($\mathcal{B}$) becomes admissible for ranking in addition to effects No. 4 and 6 at the second step.

Note 2.6. The last 5 lines of the I(X)-printouts are the first five of FEMO, single dropping. Note that the last I(X)-value is smaller than ALPHA(1) = 0.05. Therefore, a full printout is given for step No. 5 of FEMO, single dropping.

Note 2.7. The FCA for FEMO, cumulative dropping, shows that there are no significant effects. However, one must not forget that this conclusion is based on only 1 degree of freedom for error. For the orthogonal data layout of this example, the DIFF MS - column shows the mean squares as given by Davies [1956], p. 456.

Note 2.8. The FCA of FEMO, single dropping, does show a significant model which contains the two main effects $\mathcal{A}$ and $\mathcal{B}$, with $\mathcal{B}$ being the effect ranked most important. By covious reasons in the present example, which is exhibited for the purpose mentioned, it does not make sense to try to close the gap between the results of the two ranking procedures.

Note 2.9. Problem running time was 24 seconds. This is relatively long; however, one must not forget that a program like NOVACOM is bound to be inefficient (timewise) for a small case as the present one.


DATA CARD LAYOUT
Ident C
C C 1
C C 2
C C 3 (Blank)
C C 4
1st Data C
2nd Data C
1st Data C
2nd Data C
1st Data C
2nd Data C
EXAMPLE 2 HALF REPLICATE OF 2x2x2x2 FACTORIAL (DAVIES, 1956)
LAST

```
NOVACOM ... NON-ORTHOGONAL VARIANCE AND COVARIANCE ANALYSIS BY MULTIPLE REGRESSION ... PROGRAM VERSION OF JULY 19,1967

 E NOSODV  D  GDD  NCC4 NSCC4  TP  P GDC NCC5 CODE    C    DROP  KALPHA ALPHA(1) ALPHA(2) ALPHA(3)  CAD    TOLI2    IRCO ADA
 4    1    1   6     2    1     0  0       0    1    1.00    1      1    .0500    .0000    .0000    0   .1000E-03    1   0
  1  1  1  1  2  2  2  1  3  3  3  1  4  4  4  1  5  5  1  1  2  1  6  6  1  1  3  1
EXAMPLE 2... HALF REPLICATE OF 2X2X2X2 FACTORIAL (DAVIES,1956)
/    /  /      /SET  1

IDENTIFICATION OF INDEPENDENT VARIABLES (IV-S)

DIV-S...

        EFFECT NUMBER  1

        IV   1      1* 1

        EFFECT NUMBER  2

        IV   2      2* 1

        EFFECT NUMBER  3

        IV   3      3* 1

        EFFECT NUMBER  4

        IV   4      4* 1

        EFFECT NUMBER  5

        IV   5      1* 1X 2* 1

        EFFECT NUMBER  6

        IV   6      1* 1X 3* 1
```

Note 2.1

Note 2.2

EXAMPLE 2... HALF REPLICATE OF 2X2X2X2 FACTORIAL (DAVIES,1956)
/ / / /SET 1

Note 2 3

FULL DATA MATRIX

CELL IDENTIFICATION	(1) (V1)	(2)	(3)	(4)	(5)	(6)
1	.100000E+01	.100000E+01	.100000E+01	.100000E+01	.100000E+01	.100000E+01
	127.000000000					
	1, 1, 1, 1					
2	.100000E+01	.100000E+01	.000000E 00	.000000E 00	.100000E+01	.000000E 00
	126.000000000					
	1, 1, 2, 2					
3	.100000E+01	.000000E 00	.000000E 00	.100000E+01	.000000E 00	.000000E 00
	120.000000000					
	1, 2, 2, 1					
4	.100000E+01	.000000E 00	.100000E+01	.000000E 00	.000000E 00	.100000E+01
	122.000000000					
	1, 2, 1, 2					
5	.000000E 00	.100000E+01	.000000E 00	.100000E+01	.000000E 00	.000000E 00
	121.000000000					
	2, 1, 2, 1					
6	.000000E 00	.100000E+01	.100000E+01	.000000E 00	.000000E 00	.000000E 00
	114.000000000					
	2, 1, 1, 2					
7	.000000E 00	.000000E 00	.100000E+01	.100000E+01	.000000E 00	.000000E 00
	130.000000000					
	2, 2, 1, 1					
8	.000000E 00	.000000E 00	.000000E 00	.000000E 00	.000000E 00	.000000E 00
	132.000000000					
	2, 2, 2, 2					

Follows: Summation matrix
Averages
Total Sum of Squares adjusted for the mean

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  4/  5/  6/
```

Note 2.4

```
REGRESSION  COEFFICIENTS,STEP   1 /FEMO/Y1/CUMUL /SET  1
(  0)  0.13275000000000E+03 (  1)  -.13000000000000E+02 (  2)  -.13500000000000E+02 (  3)  -.45000000000000E+01
(  4)  0.10000000000000E+01 (  5)  -.10000000000000E+01 (  6)  0.60000000000000E+01
```

Follows: Full Printout of Step Number 1, FEMO, Cumulative dropping

```
ARG1=   0.50  ARG2=   0.50  X= .69230769E+00  I(X)= 0.82966591637800E+00
ARG1=   0.50  ARG2=   0.50  X= .90000000E+00  I(X)= 0.79516723530087E+00
ARG1=   0.50  ARG2=   0.50  X= .20000000E+00  I(X)= 0.29516723530086E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/  4/  6/
```

Note 2.5

```
REGRESSION  COEFFICIENTS,STEP   2 /FEMO/Y1/CUMUL /SET  1
(  0)  0.13300000000000E+03 (  1)  -.13500000000000E+02 (  2)  -.14000000000000E+02 (  3)  -.45000000000000E+01
(  4)  0.10000000000000E+01 (  6)  0.60000000000000E+01
ARG1=   0.50  ARG2=   1.00  X= .11335013E-01  I(X)= 0.10646601614815E+00
ARG1=   0.50  ARG2=   1.00  X= .64285714E+00  I(X)= 0.80179372573727E+00
ARG1=   0.50  ARG2=   1.00  X= .19565217E+00  I(X)= 0.44232588846469E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/  6/

REGRESSION  COEFFICIENTS,STEP   3 /FEMO/Y1/CUMUL /SET  1
(  0)  0.13350000000000E+03 (  1)  -.13500000000000E+02 (  2)  -.14000000000000E+02 (  3)  -.45000000000000E+01
(  6)  0.60000000000000E+01
ARG1=   0.50  ARG2=   1.50  X= .11278195E-01  I(X)= 0.13496197569645E+00
ARG1=   0.50  ARG2=   1.50  X= .18000000E-00  I(X)= 0.52351566578212E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/  2/  3/

REGRESSION  COEFFICIENTS,STEP   4 /FEMO/Y1/CUMUL /SET  1
(  0)  0.13200000000000E+03 (  1)  -.10500000000000E+02 (  2)  -.14000000000000E+02 (  3)  -.15000000000000E+01
ARG1=   0.50  ARG2=   2.00  X= .18329939E-01  I(X)= 0.20184132928925E+00
ARG1=   0.50  ARG2=   2.00  X= .10791367E-01  I(X)= 0.15524174452641E+00
ARG1=   0.50  ARG2=   2.00  X= .15254237E-00  I(X)= 0.55606111130792E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/  2/

REGRESSION  COEFFICIENTS,STEP   5 /FEMO/Y1/CUMUL /SET  1
(  0)  0.13125000000000E+03 (  1)  -.13500000000000E+02 (  2)  -.14000000000000E+02
ARG1=   0.50  ARG2=   2.50  X= .18000000E-01  I(X)= 0.22571968435672E+00
ARG1=   0.50  ARG2=   2.50  X= .10676157E-01  I(X)= 0.17447594217656E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/

REGRESSION  COEFFICIENTS,STEP   6 /FEMO/Y1/CUMUL /SET  1
(  0)  0.12600000000000E+03 (  2)  -.14000000000000E+02
ARG1=   0.50  ARG2=   3.00  X= .70083458E-02  I(X)= 0.15624644055430E+00
ARG1=   0.50  ARG2=   0.50  X= .90000000E+00  I(X)= 0.79516723530087E+00
ARG1=   1.00  ARG2=   0.50  X= .71428571E+00  I(X)= 0.46547751617515E+00
ARG1=   1.50  ARG2=   0.50  X= .28000000E-00  I(X)= 0.69136869268428E-01
ARG1=   2.00  ARG2=   0.50  X= .34745763E+00  I(X)= 0.44393888869207E+00
ARG1=   2.50  ARG2=   0.50  X= .11800000E-00  I(X)= 0.16975580279674E-02
```

Note 2.6

Follows: Full Printout of Step Number 5, FEMO, Single dropping

EXAMPLE 2... HALF REPLICATE OF 2X2X2X2 FACTORIAL (DAVIES,1956)
/FE=B/F1/CUMUL /SET A1

FINAL COMPREHENSIVE FEMO-CUMULATIVE DOPPING Note 2.7

STEP	EFFECT	PRC	F(X)	DIFF MS	DIF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	0001*0002		0.78510724	.50000000	001	.11111111	.50000000	001	4.5000000	001	.99299065
002	0004		0.48547752	0000000002	00_1	.27777778	1.2500000	00_2	4.5000000	00_1	.99221184
003	0001*0003		0.48235167	0000000018	001	1.5185185	6.8333333	003	4.5000000	001	.98809657
004	0003		0.45513111	4.5000000	00_1	1.3888889	6.2500000	00_4	4.5000000	00_1	.98105919
005	0001		0.22571948	220.50000	001	10.011111	49.100000	005	4.5000000	001	.95404984
006	0002		0.15624444	0000000392	00_1	23.611111	106.25000	006	4.5000000	00_1	.61099190

EXAMPLE 2... HALF REPLICATE OF 2X2X2X2 FACTORIAL (DAVIES,1956)
/FE=B/F1/SINGLE/SET A1

FINAL COMPREHENSIVE FEMO-SINGLE DOPPING Note 2.8

STEP	EFFECT	PRC	F(X)	DIFF MS	DIF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	0001*0002		0.78510724	.50000000	001	.11111111	.50000000	001	4.5000000	001	.99299065
002	0004		0.48547752	0000000002	00_1	.80000000	0000000002	00_1	2.5000000	002	.99221184
003	0001*0003		0.06913697	0000000018	001	7.7142857	0000000018	001	2.3333333	003	.98809657
004	0003		0.44393849	4.5000000	00_1	.72000000	4.5000000	00_1	6.2500000	004	.98105919
005	0001	*	0.00168758	220.50000	001	37.372881	220.50000	001	5.9000000	005	.95404984
006	0002		0.02201958	0000000392	00_1	9.4000000	0000000392	00_1	41.666667	006	.61099190

PROBLEM RUNNING TIME(HRS,MIN,SEC)=00000624 Note 2.9

3.4.3 Example 3

Example 3 is exhibited in order to illustrate the effect of coding and of the restricted admissibility rules upon the determination of a significant model when there are quantitative factors. At the same time, the example demonstrates the applicability of NOVACOM in analyzing orthogonal date layouts when the quantitative factor levels are non-equidistant.

		β_1 $X_{b1}=0$	β_2 $X_{b2}=1$	β_3 $X_{b3}=3$	β_4 $X_{b4}=6$
α_1	$X_{a1}=0$	16 14	12 11	17 19	13 11
		$\Sigma=30$	$\Sigma=23$	$\Sigma=36$	$\Sigma=24$
α_2	$X_{a2}=2$	15 15	14 17	15 18	12 14
		$\Sigma=30$	$\Sigma=31$	$\Sigma=33$	$\Sigma=26$
α_3	$X_{a3}=5$	10 9	7 6	10 14	9 13
		$\Sigma=19$	$\Sigma=13$	$\Sigma=24$	$\Sigma=22$

Table 3.2

Data Layout Example 3

The data as exhibited in Table 3.2 is taken from two sources: the 24 values of the response variable, y, are from Hicks [1964], p. 129, and the quantitative factor variable values are from Robson [1959]. (The totals for each cell are given for purposes to be seen later.)

The data in the 3x4 classification with both factors α and β being quantitative (leading to a breakdown into 11 one-degree-of-freedom components of the sum of squares between cells) is treated in three different ways:

A (Example 3A): factor levels uncoded; restricted admissibility in the ranking of factorial effects.

B (Example 3B): factor levels uncoded; unrestricted admissibility in the ranking of factorial effects.

C (Example 3C): factor levels coded; unrestricted admissibility in the ranking of factorial effects.

The resulting significant models are different in all three cases as will be discussed with the reproduced printout of the FCAs. Besides the FCAs again only those parts of the actual printout are reproduced which show features not exhibited in the two previous examples.

Notes on printout Example 3A.

Note 3.1. The level values (values of the quantitative factor variables) are identified.

Note 3.2. The quantitative factorial effects (all with 1 degree of freedom) are identified.

Note 3.3. Only effect No. 11 ($\alpha_{quadr.} \times \beta_{cubic}$) is admissible for ranking at the first step since all other effects are sub-effects of this effect.

Note 3.4. The I(X)-value corresponding to effect No. 5 (β_{cubic}) is the larger one among the two computed at this step so that this effect is being dropped from the model at this step (No. 8). Since the I(X)-value corresponding to β_{cubic} is also smaller than ALPHA(2) = 0.01, a full printout for this step is given.

Note 3.5. Step No. 7 in "single dropping" yields an I(X)-value which is smaller than ALPHA(2) = 0.01. Since no full printout for step No. 7 had been given before, it is given here.

Note 3.6. With restricted admissibility in the FEMO ranking, the "DIFF MS" column in the FCA shows, for this orthogonal case, the breakdown of the sum of squares between cells into orthogonal components as one would obtain them by application of orthogonal polynomials.

The value DIFF MS = 5.2457707 for $\alpha_{quadr.} \times \beta_{linear}$ (Step 5 of the ranking) may be checked in employing the coefficients given by Robson [1959], in his "Table 4." The sum of squares (1 degree of freedom) due to the component $Q_A L_B$ is computed as follows, using the totals from Table 3.2 given earlier:

-15·30
+25·30
-10·19
- 9·23
+15·31
- 6·13
+ 3·36
- 5·33
+ 2·24
+21·24
-35·26
+14·22

Sum = 183

Sum of squared coefficients = 3192

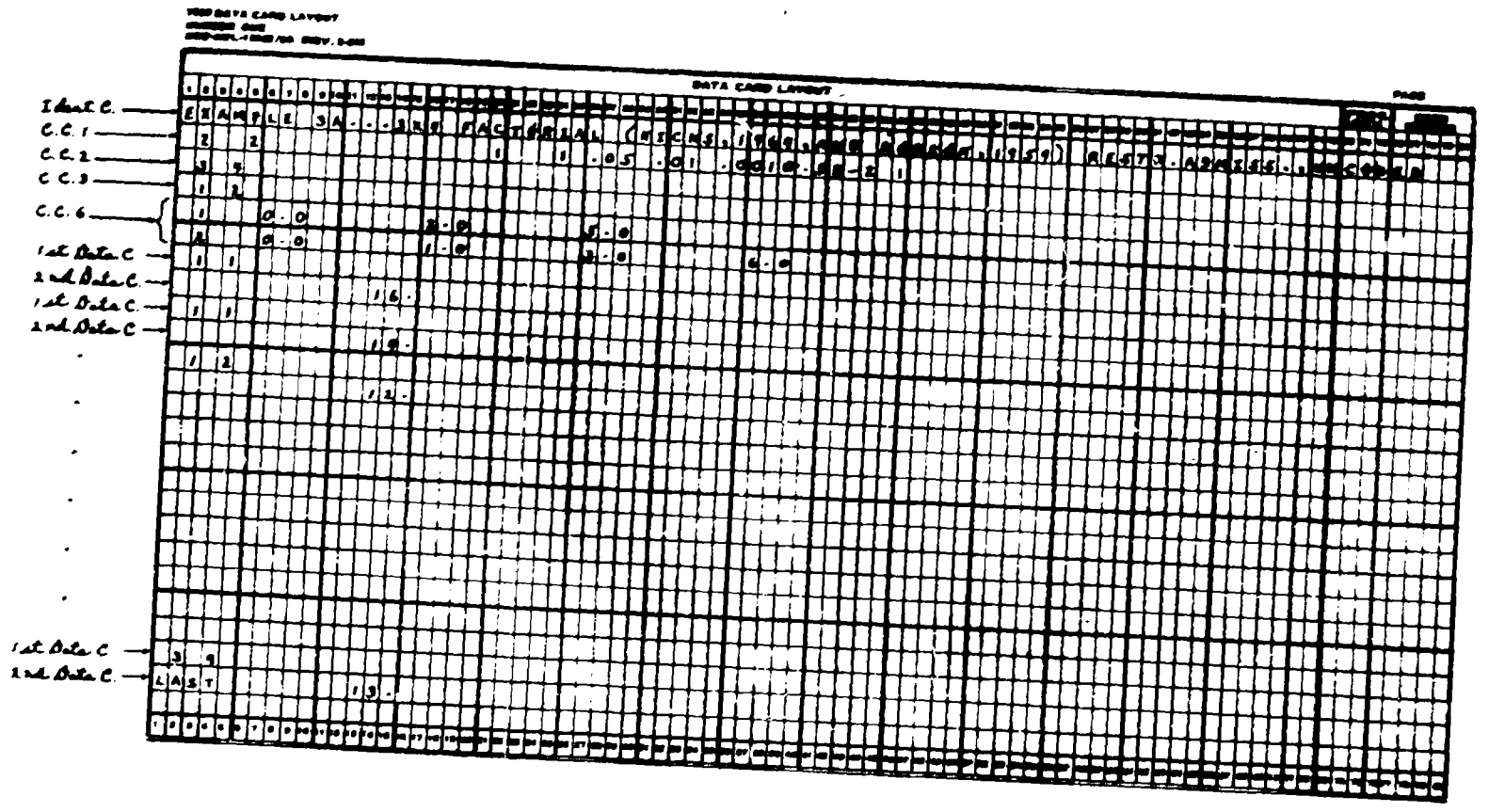
Sum of squares due to "$Q_4 L_6$" = $\frac{(183)^2}{2 \cdot 3192} = 5.2457707$.

Note that the significant model, at the $\alpha = 0.05$ level, contains all 11 effects.

Note 3.7. The FCA for the single dropping procedure shows the same significant model at the 0.05 level as did the FCA for cumulative dropping. For the 0.01 significance level there is a gap between the two models. This gap may be closed by dividing DIFF MS = 36.266447 of Step 7 in cumulative dropping by MS(2) = 2.875 to give a value of F = 12.614 which, with 1 and 12 degrees of freedom, is significant at ALPHA(2) = 0.01.

Note 3.8. This is the FCA, FEMO, cumulative dropping, for Example 3B. Due to the unrestricted admissibility in the ranking, the ranking order is different from that obtained in Example 3A. Note the differences also in the values of DIFF MS in this orthogonal case: none of the DIFF MS-values is equal for Example 3A and 3B. The significant model is equal for cumulative and single dropping at both levels ALPHA(1) = 0.05 and ALPHA(2) = 0.01. Note that at the 0.01 level of significance, the significant model as defined by the ranking now contains the effect $\mathcal{A}_{quadr.}$ only. However, dividing DIFF MS = 39.272801 for $\mathcal{B}_{quadr.}$ (Step 9) by MS(2) = 2.875 yields F = 13.660 which is also significant at the 0.01 level.

Note 3.9. Example 3C differs from Example 3B only by the coding of the quantitative factor level values. For instance, the three levels of factor $\mathcal{A}$, i.e., $X_{a1} = 0$, $X_{a2} = 2$, and $X_{a3} = 5$, with Range = 5-0 = 5, and with average = (0 + 2 + 5)/3 = 2.33, are in coded form: -0.4667, -0.0667, +0.5333. The ranking is again different (from those of Examples 3B and 3A) and the significant models contain even fewer terms than in Example 3B. The gap between the significant models of cumulative and single dropping at the 0.05 level may be closed by dividing DIFF MS = 17.402285 of $\mathcal{B}_{quadr.}$ (Step 6) by MS(2) = 2.875, leading to F = 6.053, which is significant at ALPHA(1) = 0.05.


DATA CARD LAYOUT
Ident C.
C.C. 1
C.C. 2
C C. 3
C.C. 6
1st Data C.
2nd Data C.
1st Data C.
2nd Data C.
1st Data C.
2nd Data C.
EXAMPLE SA
LAST

```
NPVACP ... N2N-ORTHOGONAL VARIANCE AND COVARIANCE ANALYSIS BY MULTIPLE REGRESSION ... PROGRAM VERSION OF JULY 15,1967

 E NPSPDV  I  JJJ  NCC4 NSCC4  TP  P GDC NCC5 CBTF    C    DRMP  KALPHA ALPHA(1) ALPHA(2) ALPHA(3)  CAD    TPLY2    IRCB ADA
 2    1    2         0    1    0    1    1    1.00    1      1    .0500    .0100    .0010    C   .5000E-02    1   0

QUANTITATIVE FACTOR LEVEL VALUES
FACTOR ...

    1        .0000000000E 00 0.2000000000E+ 1 0.5000000000E+01

    2        .0000000000E 00 0.1000000000E+ 1 0.3000000000E+01 0.4000000000E+01
  1  1 -1  1  1  2 -1  2  2  3 -2  1  2  4 -2  2  2  5 -2  3  3  6 -1  1 -2  1  3  7 -1  1 -2  2  3  8 -1  1 -2  3  3  9 -1  2 -2
  1  1 10 -1  2 -2  2  3 11 -1  2 -2  3
FXAMP E 3A,..3X4 FACTORIAL (HICKS,1964,AND RRBSON,1959) RESTR.ADMISS..UNCODED
/    /  /       /SET  1

IDENTIFICATION OF INDEPENDENT VARIABLES (IV-S)

DIV-S...

        EFFECT NUMBER  1
        IV   1      1. 1
        EFFECT NUMBER  2
        IV   2      1. 2
        EFFECT NUMBER  3
        IV   3      2. 1
        EFFECT NUMBER  4
        IV   4      2. 2
        EFFECT NUMBER  5
        IV   5      2. 3
        EFFECT NUMBER  6
        IV   6      1. 1X 2. 1
        EFFECT NUMBER  7
        IV   7      1. 1X 2. 2
        EFFECT NUMBER  8
        IV   8      1. 1X 2. 3
        EFFECT NUMBER  9
        IV   9      1. 2X 2. 1
        EFFECT NUMBER 10
        IV  10      1. 2X 2. 2
        EFFECT NUMBER 11
        IV  11      1. 2X 2. 3
```

Note 3.1

Note 3.2

Follows Full data matrix
Summation matrix
Averages, ATSS

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../ 11/ Note 3.3

REGRESSION  COEFFICIENTS,STEP   1 /FEMO/Y1/CUMUL /SET  1
(  0)  0.14999999999820E-02 (  1)  0.73333333372744E-00 (  2)  -.36666666674347E-00 (  3)  -.73999999990295E-01
(  4)  0.44499999995045E-01 (  5)  -.54999999994339E-00 (  6)  0.62822222201004E-01 (  7)  -.34105555544725E-01
(  8)  0.39499999957628E-00 (  9)  -.12077777773648E-01 ( 10)  0.64833333312258E-00 ( 11)  -.73888888844816E-01
```

Follows: Full printout for Step Number 1, FEMO, Cumulative dropping

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  8/ 10/

REGRESSION  COEFFICIENTS,STEP   2 /FEMO/Y1/CUMUL /SET  1
(  0)  0.14403409090888E-02 (  1)  0.20325757575814E-01 (  2)  -.61856060606073E-00 (  3)  -.43693181817099E-01
(  4)  0.29147727272193E-01 (  5)  -.37499999999399E-00 (  6)  -.31792929295944E-00 (  7)  -.67171717150105E-01
(  8)  0.13888888887448E-01 (  9)  0.71843434344373E-01 ( 10)  0.12626262608462E-03
ARG1=   6.00  ARG2=   1.00  X= .68088712E+00  I(X)= 0.99643895508553E-01
ARG1=   6.00  ARG2=   1.00  X= .68494020E+00  I(X)= 0.10325613002626E-00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  8/  9/

REGRESSION  COEFFICIENTS,STEP   3 /FEMO/Y1/CUMUL /SET  1
(  0)  0.14402255639075E-02 (  1)  0.20350877193025E-01 (  2)  -.61904761904732E-00 (  3)  -.43674812028915E-01
(  4)  0.29144736841927E-01 (  5)  -.37499999999347E-00 (  6)  -.32192982458924E-00 (  7)  -.66520467822189E-01
(  8)  0.13858888887296E-01 (  9)  0.72619047618968E-01
ARG1=   6.00  ARG2=   1.50  X= .68088654E+00  I(X)= 0.18767920120535E-00
ARG1=   6.00  ARG2=   1.50  X= .62033477E+00  I(X)= 0.11400312369423E-00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  5/  7/  9/

REGRESSION  COEFFICIENTS,STEP   4 /FEMO/Y1/CUMUL /SET  1
(  0)  0.14291775841120E-02 (  1)  0.20824362041451E-01 (  2)  -.61904761904742E-00 (  3)  -.38062438292679E-01
(  4)  0.29301723374888E-01 (  5)  -.34259259259226E-00 (  6)  -.56246012759041E-00 (  7)  0.55322966507058E-01
(  8)  0.72619047618949E-01
ARG1=   6.00  ARG2=   2.00  X= .36748016E+00  I(X)= 0.11808719264374E-01
ARG1=   6.00  ARG2=   2.00  X= .62663976E+00  I(X)= 0.19618885184211E-00
ARG1=   6.00  ARG2=   2.00  X= .61700790E+00  I(X)= 0.18196537514041E-00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  5/  9/

REGRESSION  COEFFICIENTS,STEP   5 /FEMO/Y1/CUMUL /SET  1
(  0)  0.14789682539684E-02 (  1)  0.19690476190458E-01 (  2)  -.61904761904730E-00 (  3)  -.65992063492020E-01
(  4)  0.27592592592502E-01 (  5)  -.34259259259221E-00 (  6)  -.22261904761831E-00 (  9)  0.72619047618918E-01
ARG1=   6.00  ARG2=   2.50  X= .35107736E+00  I(X)= 0.16071261804100E-01
ARG1=   6.00  ARG2=   2.50  X= .57212672E+00  I(X)= 0.18840080349565E-00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/  5/  6/

REGRESSION  COEFFICIENTS,STEP   6 /FEMO/Y1/CUMUL /SET  1
(  0)  0.15219663742690E-02 (  1)  0.93264411027534E-00 (  2)  -.43749999999995E-00 (  3)  -.47711908304041E-01
(  4)  0.27592592592580E-01 (  5)  -.34259259259219E-00 (  6)  0.15194239588979E-00
ARG1=   6.00  ARG2=   3.00  X= .35726204E+00  I(X)= 0.28136099177998E-01
ARG1=   6.00  ARG2=   3.00  X= .33328596E+00  I(X)= 0.19647066107138E-01
ARG1=   6.00  ARG2=   3.00  X= .47531573E+00  I(X)= 0.11450151959313E-00
```

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/  5/

REGRESSION  COEFFICIENTS,STEP   7 /FEMO/Y1/CUMUL /SET  1
(  0)  0.1433333333333E+02 (  1)  0.1512409999999997E-01 (  2)  -.4374999999999995E-00 (  3)  -.4414404644644607E+01
(  4)  0.2759259259259578E+01 (  5)  -.3425925925925718E-00
ARG1=   6.00  ARG2=   3.50  X= .9169959?E-00  I(X)= 0.2306055180412185-01
ARG1=   6.00  ARG2=   3.50  X= .2979335?1E-00  I(X)= 0.1876099513713?E-01
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/  5/

REGRESSION  COEFFICIENTS,STEP   8 /FEMO/Y1/CUMUL /SET  1
(  0)  0.1536951794380?E+02 (  1)  -.9440789473684?9E-00 (  3)  -.4416666666662?E-01 (  4)  0.2759259259259?6?E-01
(  5)  -.3425925?9?9?23E-00
ARG1=   6.00  ARG2=   4.00  X= -.17322178E-00  I(X)= 0.1403842058?6?13E-02
ARG1=   6.00  ARG2=   4.00  X= .2268??40E-00  I(X)= 0.6?1336132?99??9E-02
```

Note 3.4

Follows: Full Printout for Step Number 8, FEMO, Cumulative dropping

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/  4/

REGRESSION  COEFFICIENTS,STEP   9 /FEMO/Y1/CUMUL /SET  1
(  0)  0.1428158825?9?31E+02 (  1)  -.9440789473684?9E-00 (  3)  0.1516441414141326E-01 (  4)  -.2462121212119?6E-00
ARG1=   6.00  ARG2=   4.50  X= .1423384?5E-00  I(X)= 0.?144?35?47?4?05E-03
ARG1=   6.00  ARG2=   4.50  X= .1?983843E-00  I(X)= 0.4482944519?9?59E-02
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/  3/

REGRESSION  COEFFICIENTS,STEP  10 /FEMO/Y1/CUMUL /SET  1
(  0)  0.1531?2855?5?4?6E+02 (  1)  -.9440789473684?9E-00 (  3)  0.3968253968?8?9E-02
ARG1=   6.00  ARG2=   5.00  X= -.1312?048E-00  I(X)= 0.6857?8?004?94?8E-03
ARG1=   6.00  ARG2=   5.00  X= .1?983614E-00  I(X)= 0.6342364693?9?58E-02
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/

REGRESSION  COEFFICIENTS,STEP  11 /FEMO/Y1/CUMUL /SET  1
(  0)  0.1531?184?1?8?26E+02 (  1)  -.9440789473684?9E-00
ARG1=   6.00  ARG2=   5.50  X= .1311?949E-00  I(X)= 0.943947398?19?60E-03
```

Follows: Full Printout for Step Number 11, FEMO, Cumulative dropping

```
ARG1=   6.00  ARG2=   0.50  X= .6849487?8E+00  I(X)= 0.3675153?78957?56-0?
ARG1=   6.50  ARG2=   0.50  X= .9999991?E+00  I(X)= 0.9974?1?2?3?2??9E-00
ARG1=   7.00  ARG2=   0.50  X= .?9408?74E+00  I(X)= 0.?77?8?99?6?54?35E-00
ARG1=   7.50  ARG2=   0.50  X= .9?0329?9E+00  I(X)= 0.?72?99?00487618-00
ARG1=   8.00  ARG2=   0.50  X= .?1?08?37?E+00  I(X)= 0.?3476?5?5?9?543E-00
ARG1=   8.50  ARG2=   0.50  X= .83C?87?0E+00  I(X)= 0.?8?16179?4?4?00E-0?
ARG1=   9.00  ARG2=   0.50  X= .68682113E+00  I(X)= 0.?77?2?54?49?4?02E-02
```

Note 3.5

Follows: Full Printout for Step Number 7, FEMO, Single dropping, and removing 4 lines of I(X) - information

EXAMPLE 5A...3K4 FACTORIAL (HICKS,1964,AND BABSON,1959) TESTR.ADMISS,,UNCODED
/FEMD/V1/SINGLE/SET 01

FINAL COMPREHENSIVE FEMD-SINGLE DROPPING

Note 3.6

STEP	EFFECT	PRC	TEST	DIFF MS	DIF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	1.02AB2.33	*	0.03675154	15.869318	001	5.9197628	15.869318	001	2.8750000	012	.8888CC91
002	1.02A 2.32		0.10329615	.00004272	001	2.7598889	7.9344804	002	2.8750000	012	.80845133
003	1.01A 2.33		0.18767920	.29987374	001	1.9744939	5.3897449	003	2.8750000	012	.80845117
004	1.01A2.32		0.19818885	4.3861209	001	1.7876398	5.1388889	004	2.8750000	012	.80731079
005	1.02A 2.31		0.18660860	5.2497707	001	1.7968749	5.1402552	005	2.8750000	012	.79083012
006	1.01A2.31		0.11450132	12.283807	001	2.2877295	6.3472222	006	2.8750000	012	.77068106
007	1.02		0.02350855	36.266447	001	3.6043000	10.621397	007	2.8750000	012	.7239768$_1$
008	2.03	*	0.00601836	43.213384	001	5.1114419	14.693306	008	2.8750000	012	.3868$4693
009	2.02		0.00440294	20.976299	001	5.3287332	15.348629	009	2.8750000	012	.6811721^{45}
010	2.01		0.00034234	.00198413	001	4.8849199	13.814165	010	2.8750000	012	.34347217
011	21.01	*	0.00004305	90.316884	001	7.2230789	20.768939	011	2.8750000	012	.3434$463

EXAMPLE 5A...3K4 FACTORIAL (HICKS,1964,AND BABSON,1959) TESTR.ADMISS,,UNCODED
/FEMD/V1/SINGLE/SET 01

FINAL COMPREHENSIVE FEMD-SINGLE DROPPING

Note 3.7

STEP	EFFECT	PRC	TEST	DIFF MS	DIF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	1.02AB2.33	*	0.03675154	15.869318	001	5.9197628	15.869318	001	2.8750000	012	.8888CC91
002	1.02A 2.32		0.00740183	.00004272	001	.0000$183	.00004272	001	3.8749679	013	.80845133
003	1.01A 2.33		0.77733991	.29987374	001	.00134693	.29987374	001	3.5978113	014	.80845117
004	1.01A 2.32		0.27239520	4.3861209	001	1.2989168	4.3861209	001	3.3778498	015	.80731079
005	1.02A 2.31		0.23478979	5.2497707	001	1.5000000	5.2497707	001	3.4499722	016	.79083012
004	1.01A 2.31		0.08018171	12.283807	001	3.4659179	12.283807	001	3.5471368	017	.77068106
007	1.02	*	0.00770285	36.266447	001	8.9937698	36.266447	001	4.8324874	018	.7239768$_1$
008	2.03		0.01203326	43.213384	001	7.5430036	43.213384	001	5.7200358	019	.3868$4693
009	2.02		0.11950685	20.976299	001	2.7062891	20.976299	001	7.6831502	020	.6811721^{45}
010	2.01		0.00775164	.00198413	001	.00274139	.00198413	001	8.2700268	021	.34347217
011	21.01		0.00281757	90.316884	001	11.900838	90.316884	001	7.8473305	022	.3434$463

PROBLEM SUMMARY TIMES=MPS,MIN,SEC=47C0C0037

EXAMPLE 38...1964 FACTORIAL (MICHS,1964,AND BRESN..1959) UNDER'0.AD=ISS..UNCBOEP
/FEMD/Y1/COMPUL /SET 01

FINAL COMPREHENSIVE FEMD-CUMULATIVE GROUPING

STEP	EFFECT	PRC	F(1)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	Y1.01		0.50721381	.80463956	001	.31117898	.80463956	001	2.8750000	012	.8688009[illegible]
002	Y1.02X02.03	*	0.84113422	23.329269	001	4.2171545	13.109944	002	3.8750000	012	.8693e83e
003	Y1.02X02.[illegible]		0.[illegible]3863	.82761836	001	2.9040985	8.3491682	001	2.8750000	012	.77669500
004	Y1.01X02.02		0.11247283	2.0152352	001	2.3532817	6.7056849	004	2.8750000	012	.77354770
005	Y1.01X02.01		0.14123890	2.9822651	001	2.0672612	5.9290810	005	2.8750000	012	.76988409
006	Y1.02X01.01		0.18857119	.86786486	001	1.7688984	5.0054449	006	2.8750000	012	.7568[illegible]00
007	Y1.02X02.02		0.00181078	19.033811	001	2.4619369	7.0780686	007	2.8750000	012	.75278437
008	Y2.01		0.01[illegible]110	20.892007	001	3.0625429	8.8048109	008	2.8750000	012	.68038[illegible]38
009	Y2.02		0.01143262	38.272801	001	4.2488498	12.190143	009	2.8750000	012	.60009112
010	Y2.01		0.0[illegible]400255	3.4266.10	001	3.9421984	11.3338[illegible]5	010	2.8750000	012	.45[illegible]94[illegible]22
011	Y1.02	*	0.00004395	119.120[illegible]0	001	7.2239780	20.768[illegible]39	011	2.8750000	012	.43778878

EXAMPLE 38...1964 FACTORIAL (MICHS,1964,AND BRESN..1959) UNDER'0.AD=ISS..UNCBOEP
/FEMD/Y1/SINGLE/SET 01

FINAL COMPREHENSIVE FEMD-SINGLE GROUPING

STEP	EFFECT	PRC	F(1)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	Y1.01		0.50721381	.80463956	001	.31117898	.80463956	001	2.8750000	012	.8688009[illegible]
002	Y1.02X02.03	*	0.8[illegible]178257	23.329269	001	8.9678668	23.329269	001	2.7228668	013	.96519210
003	Y1.01X02.03		0.66168166	.82761836	001	.19738098	.82761836	001	4.1942777	014	.77669500
004	Y1.01X02.02		0.48[illegible]09933	2.0152352	001	.5878319[illegible]9	2.[illegible]152352	001	3.8490336	015	.77354770
005	Y1.01X02.01		0.42468906	2.9822651	001	.67112618	2.5022651	001	3.8478712	016	.76988409
006	Y1.02X02.01		0.65767812	.86786486	001	.22998247	.86786486	001	3.7732356	017	.75684400
007	Y1.02X02.02		0.05392743	19.033811	001	5.2498742	19.033811	001	3.8118158	018	.75278437
008	Y2.01		0.04781458	20.892007	001	4.7279597	20.892007	001	4.4234998	019	.68038100
009	Y2.02		0.01273623	38.272801	001	7.4449184	38.272801	001	5.2469744	020	.60009112
010	Y2.01		0.[illegible]541273	3.4266569	001	.5281[illegible]96	3.4266569	001	6.8672062	021	.45[illegible]99122
011	Y1.02	*	0.00042969	119.120[illegible]0	001	17.131193	119.120[illegible]0	001	8.7[illegible]9158	022	.43778878

EXAMPLE 3C...3X4 FACTORIAL (WICKS,1964,AND RUBSON,1959) UNRESTR.ADMISS..CODED
/FEMB/Y1/CUMUL /SET 01

FINAL COMPREHENSIVE FEMB-CUMULATIVE DROPPING

Note 3.4

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	[illegible]		0.91064700	.03774966	001	.01313032	.03774966	001	2.8750000	012	.86880051
002	2.01		0.97302288	.10799980	001	.02934773	.07287473	002	2.8750000	012	.86865695
003	2.03		0.99503499	.05209647	001	.02273866	.06594864	003	2.8750000	012	.86824624
004	1.01X02.01		0.83034494	3.9756920	001	.36291293	1.8433747	004	2.8750000	012	.86884812
005	1.01X02.03		0.44859972	10.474909	001	1.0109766	2.9295574	005	2.8750000	012	.05292918
006	2.02		0.17003337	17.402705	001	1.8579793	5.3416788	006	2.8750000	012	.01309667
007	1.02		0.11619333	11.282992	001	2.1531660	6.1903522	007	2.8750000	012	.74691788
008	1.02X02.02		0.11326809	5.9168859	001	2.1612761	6.1561649	008	2.8750000	012	.70401217
009	1.01	*	0.01097664	61.613531	001	4.2040000	12.310000	009	2.8750000	012	.68191094
010	1.02X02.03	*	0.00072235	113.73455	001	7.0128849	22.459743	010	2.8750000	012	.44720184
011	1.02X02.01		0.00094395	3.8609032	001	7.2239709	20.764939	011	2.8750000	012	.01460257

EXAMPLE 3C...3X4 FACTORIAL (WICKS,1964,AND RUBSON,1959) UNRESTR.ADMISS..CODED
/FEMB/Y1/SINGLE/SET 01

FINAL COMPREHENSIVE FEMB-SINGLE DROPPING

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	1.01X02.02		0.91064700	.03774966	001	.01313032	.03774966	001	2.8750000	012	.86880051
002	2.01		0.84333356	.10799980	001	.04065118	.10799980	001	2.6567500	013	.86865695
003	2.03		0.88678646	.05209647	001	.02109166	.05209647	001	2.4746964	014	.86824624
004	1.01X02.01		0.20957018	3.9756920	001	1.7186866	3.9756920	001	2.3131897	015	.86884812
005	1.01X02.03		0.05378653	10.474909	001	4.3334231	10.474909	001	2.4170937	016	.05292918
006	2.02	*	0.02523978	17.402705	001	6.8193725	17.402705	001	2.8918464	017	.01309667
007	1.02		0.09769919	11.282992	001	3.9519028	11.282992	001	3.6972261	018	.74691788
008	1.02X02.02		0.24419564	5.9168859	001	1.4443951	5.9168859	001	4.0964459	019	.70401217
009	1.01	*	0.00103241	61.613531	001	14.713793	61.613531	001	6.1874676	020	.68191094
010	1.02X02.03	*	0.00057175	113.73455	001	16.430780	113.73455	001	8.9220420	021	.44720184
011	1.02X02.01		0.57274534	3.8609032	001	.32782964	3.8609032	001	11.777154	022	.01460257

3.4.4 Example 4

Example 4 is the first in this series of examples of NOVACOM applications which deals with an incomplete and unbalanced data layout. The data is that of the example treated by Stevens [1948] and the layout of the values of the response variable y = "gain in weight" is given in Table 3.3 in slightly different arrangement than given in "Table 1" on page 349 of the Stevens paper. The factor symbols used correspond to the 3 (qualitative) factors as follows:

Factor $\mathcal{A}$: Sex ($\mathcal{A}_1$ = "M", $\mathcal{A}_2$ = "F")

Factor $\mathcal{B}$: Type of wheat in diet ($\mathcal{B}_1$ = "A", $\mathcal{B}_2$ = "B", $\mathcal{B}_3$ = "C", $\mathcal{B}_4$ = "D")

Factor $\mathcal{C}$: Litter ($\mathcal{C}_1$ = "I", $\mathcal{C}_2$ = "II", $\mathcal{C}_3$ = "III", $\mathcal{C}_4$ = "IV")

		$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$
$\mathcal{A}_1$	$\mathcal{B}_1$	43 58	73 59	81	67
	$\mathcal{B}_2$	93 83	75 89	101	100
	$\mathcal{B}_3$	91	85	92 88	106
	$\mathcal{B}_4$	89 89	98	105 108	109
$\mathcal{A}_2$	$\mathcal{B}_1$	58		62	71
	$\mathcal{B}_2$	60	71	76	
	$\mathcal{B}_3$	70	70 58		73
	$\mathcal{B}_4$		69 72		76

Table 3.3

Data Layout Example 4

The fitting of constants follows the rules given in Appendix A. All main effects and all two-factor interactions can be fitted, however, because of the five empty cells, only 4 constants for the $\mathcal{AC}$ interaction

can be fitted. This interaction would be represented by 9 constants (9 degrees of freedom) in a "full model." There are no "identities" for this layout, that is, there is no confounding among the factorial effects. The fitting process for the 4 three-factor-interaction constants is illustrated in Table 3.4 where the types of checkmarks explained in Appendix A are used. The four circled "X's" indicate the four constants fitted: abc_{111}, abc_{113}, abc_{121}, and abc_{132}. (Other sets of 4 constants could have been chosen for the $\mathcal{ABC}$ interaction.)

		C_1	C_2	C_3	C_4
A_1	B_1	(X)	✓	(X)	✓
	B_2	(X)	✓	✓	✓
	B_3	✓	(X)	✓	✓
	B_4	✓	✓	✓	✓
A_2	B_1	✓		✓	✓
	B_2	✓	✓	✓	
	B_3	✓	✓		✓
	B_4		✓		✓

Table 3.4

Fitting of $\mathcal{ABC}$ Constants in Example 4

The model containing 26 DIVs was fitted by generating automatically the full model of order D=2 (CC No. 1, columns 4-5) and by adding the four $\mathcal{ABC}$-constants via CC No. 4; see the reproduced input sheet.

Of the NOVACOM printout for Example 4 only the two FCAs are reproduced. The cumulative dropping procedure results in a significant model (at the level ALPHA(2) = 0.05) containing only the main effects of $\mathcal{A}$ and $\mathcal{B}$. The single dropping procedure results in a model, at again the 0.05 level, containing the effects $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{AB}$. Dividing

DIFT MS = 197.28205 of $\alpha\beta$ in cumulative dropping (Step 4) by MS(2) = 49.722222 yields F=3.968 which with 3 and 9 degrees of freedom is significant at α = 0.05. Therefore, the gap between the two significant models can be closed, and the conclusion would be that the significant model, at the 0.05 level, contains the three main effects (with 7 degrees of freedom) and the α x β interaction (with 3 degrees of freedom). The ranking order within the significant model shows that α and β are of approximately equal importance, whereas γ and $\alpha\beta$ are less important with $\alpha\beta$ being marginally significant.

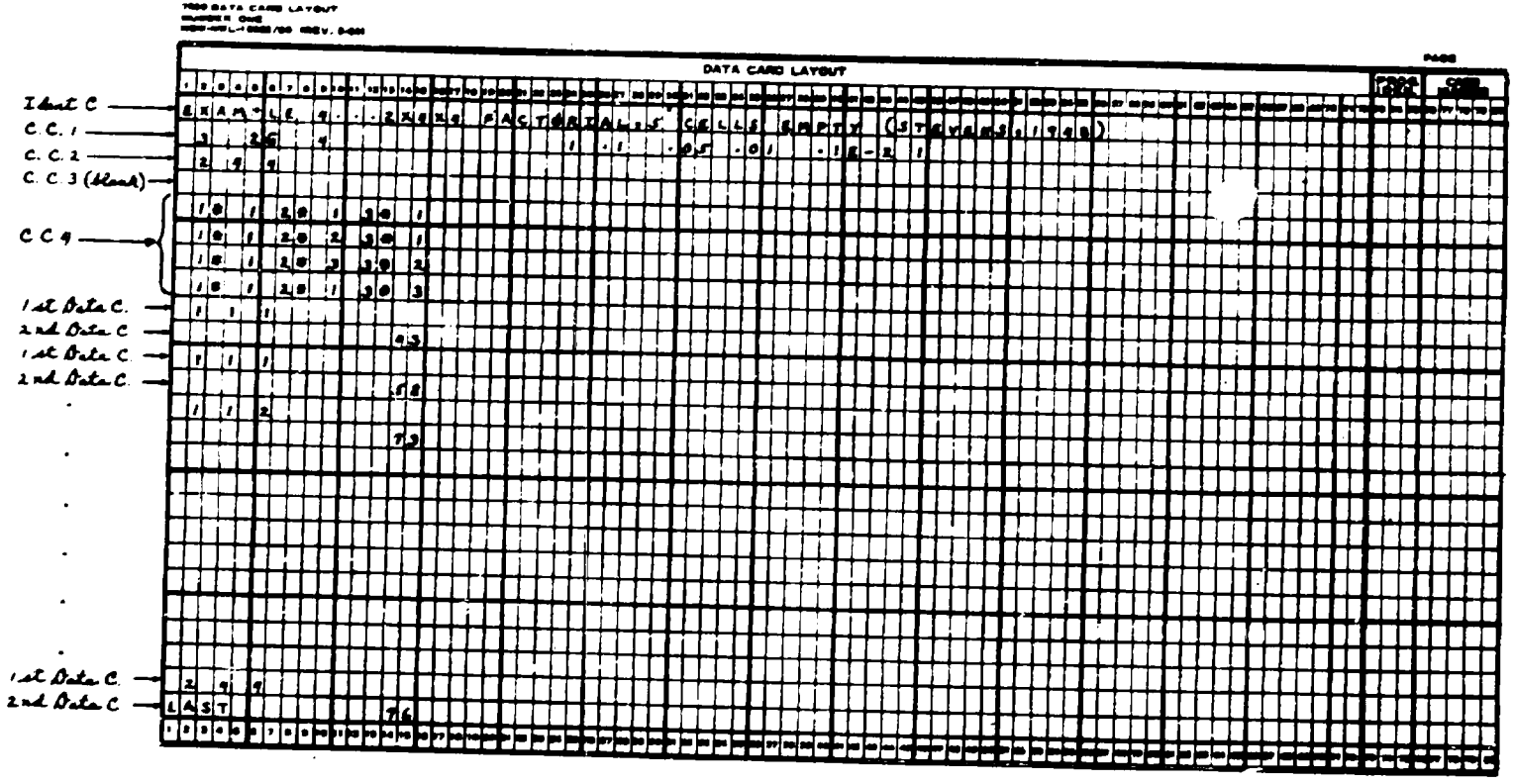
The above conclusions are essentially those which are reached in the analysis by Stevens ("Table 12" on page 365 of the paper). However, Stevens separated one degree of freedom from the $\alpha\beta$-interaction which enabled him to allocate the significance of $\alpha\beta$ to this one degree of freedom. In NOVACOM, the split-up of qualitative factorial effects into single-degree-of-freedom contrasts is not possible.

Table 12 of the Stevens paper also allows a comparison with the sums of squares obtained by NOVACOM.

The following table of values is computed from the FCA, cumulative dropping, columns "MS(1)" and "DF(1)."

Step	SS(1)=MS(1)xDF(1)	DF(1)	Due to	Stevens Value
7	9272.50	26	all effects	9272
4	1401.07	19	all interactions	--
Difference	7871.43	7	3 main effects	7871
7	9272.50	26	all effects	9272
3	809.22	16	$\alpha\gamma$,$\beta\gamma$,$\alpha\beta\gamma$	813
Difference	8463.28	10	α,β,γ,$\alpha\beta$	8459
4	1401.07	19	all interactions	--
3	809.22	16	$\alpha\gamma$,$\beta\gamma$,$\alpha\beta\gamma$	813
Difference	591.85	3	$\alpha\beta$	588

The discrepancies between the values obtained by Stevens and those by NOVACOM are small and may be attributed to the lesser accuracy of the computational procedure employed by Stevens. Note that the Stevens value 588 for the sum of squares due to $\alpha\beta$ is the sum of the values 462 and 126 in his Table 12.


DATA CARD LAYOUT
Ident C
C. C. 1
C. C. 2
C. C. 3 (blank)
C C 4
1st Data C.
2nd Data C
1st Data C.
2nd Data C.
1st Data C.
2nd Data C

EXAMPLE 4...2**4 FACTORIAL,5 CELLS EMPTY (STEVENS,1948)
/FEND/V1/CUMJ. /SET 01

FINAL COMPREHENSIVE FEND-CUMULATIVE DROPPING

STEP	EFFECT	PRC	F(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	00001X00002X00003		0.93813048	41.312805	004	.83087205	41.312805	004	49.722222	009	.95906891
002	00002X00003		0.43277276	63.728099	009	1.1429714	54.831084	013	49.722222	009	.93495975
003	00001X00003		0.51089843	23.473524	003	1.0171819	57.576563	016	49.722222	009	.87795225
004	00001X00002		0.27764054	197.28205	003	1.4850584	73.740570	019	49.722222	009	.87078734
005	00003	*	0.06339116	514.51393	003	2.6918754	133.84603	022	49.722222	009	.80081782
006	00001	*	0.00822928	2789.4429	001	5.0139908	249.10676	023	49.722222	009	.69101722
007	00002		0.00208433	1179.4815	003	7.1775398	356.63442	026	49.722222	009	.36403749

EXAMPLE 4...2**4 FACTORIAL,5 CELLS EMPTY (STEVENS,1948)
/FEND/V1/SINGLE/SET 01

FINAL COMPREHENSIVE FEND-SINGLE DROPPING

STEP	EFFECT	PRC	F(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	00001X00002X00003		0.93813048	41.312805	004	.83087205	41.312805	004	49.722222	009	.95196091
002	00002X00003		0.30301075	63.728099	009	1.3520418	63.728099	009	47.134709	013	.93495975
003	00001X00003		0.72988902	23.473524	003	.43531631	23.473524	003	53.922914	022	.87795225
004	00001X00002	*	0.02005676	197.28205	003	3.9249280	197.28205	003	50.268987	025	.87078734
005	00003	*	0.00061847	514.51393	003	7.7932583	514.51393	003	66.020387	028	.80081782
006	00001		0.00021846	2789.4429	001	25.492284	2789.4429	001	109.42299	031	.69101722
007	00002		0.00209320	1179.4815	003	6.1058184	1179.4815	003	193.17341	032	.36403749

PROBLEM RUNNING TIME(MMS,SEC)=00:00:48

3.4.5 Example 5

Example 5 is a numerical illustration of Example E (a 3x3x2 factorial) which is treated in general terms in Appendix A. Example 5 serves to illustrate the capability of NOVACOM in analyzing incomplete and unbalanced data classifications when there is confounding among the factorial effects.

The numerical data of the example has been generated according to the following model in which all effects involving factor C are absent:

$$y_{\alpha\beta\gamma\rho} = Y_{\alpha\beta\gamma} + e_{\alpha\beta\gamma\rho} = m + a_\alpha + b_\beta + ab_{\alpha\beta} + e_{\alpha\beta\gamma\rho},$$

where $e_{\alpha\beta\gamma\rho} \sim NID(0,1)$ and where, according to the structure of Example E, $\alpha = 1,2,3$; $\beta = 1,2,3$; $\gamma = 1,2$. Actually, therefore, one deals with a two-factor classification containing a dummy third factor, C. Consequently, the ranking process is expected to yield a significant model containing only the constants $\hat{a}_\alpha$, $\hat{b}_\beta$, and $\hat{ab}_{\alpha\beta}$.

In the construction of the data the following values were assigned to the model constants:

$$m = 13$$

$$a_1 = 4 \qquad b_1 = -3$$
$$a_2 = 11 \qquad b_2 = 8$$

$$ab_{11} = 5$$
$$ab_{21} = -19 \qquad ab_{22} = 3$$

With these values, the following Table 3.5 of expected cell means, $Y_{\alpha\beta\gamma}$, and actual "observations", $y_{\alpha\beta\gamma\rho} = Y_{\alpha\beta\gamma} + e_{\alpha\beta\gamma\rho}$, has been constructed, using a table of random normal deviates with $\sigma=1$. (See also Figure 5a in Appendix A.) For example, $Y_{111} = 13 + 4 - 3 + 5 = 19$, and $e_{1111} = 0.8$. Note also that repeated observations $y_{\alpha\beta\gamma\rho}$ have been included in 5 cells which will provide an estimate of $\sigma^2=1$ based on 5 degrees of freedom.

		B_1	B_2	B_3
A_1	C_1	(Y_{111} =19) y_{1111}=19.8		(Y_{131} =17) y_{1311}=19.8
	C_2	(Y_{112} =19) y_{1121}=18.2 y_{1122}=20.7		(Y_{132} =17) y_{1321}=15.7 y_{1322}=16.1
A_2	C_1	(Y_{211} =2) y_{2111}=2.1	(Y_{221} =35) y_{2211}=34.7 y_{2212}=35.3	
	C_2		(Y_{222} =35) y_{2221}=35.6	(Y_{232} =24) y_{2321}=24.0
A_3	C_1	(Y_{311} =10) y_{3111}=10.4 y_{3112}=10.8		(Y_{331} =13) y_{3311}=13.0
	C_2	(Y_{312} =10) y_{3121}=12.8	(Y_{322} =21) y_{3221}=21.6 y_{3222}=20.3	(Y_{332} =13) y_{3321}=13.4

Table 3.5

Data Layout Example 5

Because of the confounding in the given data layout, there are three possible models upon which the ranking process can be based, as is described in Appendix A. All three models have in common the following part:

$$Y^{(o)} = m + a_1x_1 + a_2x_2 + b_1x_3 + b_2x_4 + c_1x_5$$

$$+ ab_{11}x_6 + ab_{22}x_7 + ac_{11}x_8 + bc_{11}x_9 + abc_{111}x_{10}.$$

Each model, in addition to $Y^{(o)}$, contains two more constants, namely a pair from the three confounded constants ab_{21}, ac_{21}, and bc_{21}. Thus the three models are defined as follows:

$$\text{Model I} : Y^{(I)} = Y^{(0)} + ab_{21}x_{11}^{(I)} + ac_{21}x_{12}^{(I)}$$

$$\text{Model II} : Y^{(II)} = Y^{(0)} + ab_{21}x_{11}^{(II)} + bc_{21}x_{12}^{(II)}$$

$$\text{Model III}: Y^{(III)} = Y^{(0)} + ac_{21}x_{11}^{(III)} + bc_{21}x_{12}^{(III)}$$

The three models are generated by NOVACOM as follows. A third order model (D=3, columns 4-5, CC No. 1) is generated automatically from which, in each case, 5 DIVs are deleted via CC No. 4. The 3 respective sets of CC No. 4 have in common the 4 DIVs ab_{12}, abc_{121}, abc_{211}, and abc_{221}. In addition, CC 4 Set No. 1 contains the constant bc_{21}, Set No. 2 contains ac_{21}, and Set No. 3 contains ab_{21}, corresponding to the three models defined before.

Since Model III contains two constants (ac_{21} and bc_{21}) representing interactions with the dummy factor C, this model must be expected to yield improbable results because the effect appropriately measured by the constant ab_{21} is assigned to degrees of freedom associated with A and BC. Models I and II, however, should yield the proper significant model since both contain the constant ab_{21}.

The assumptions are verified by the results of the ranking processes as shown in the FCAs.

In this example, only the "Final FCA" is reproduced which combines the individual FCAs given for the three sets of CC 4, that is, for the three models.

In practice, the user of NOVACOM does not know which CC 4 Set will yield the proper significant model. However, as was discussed in Section 3.3.2, he may conclude that the significant model containing the smallest number of effects is the proper one. This model has been called "the most probable significant model."

Looking at the rankings as established for the three models (the FCAs for "SET 1", "SET 2", and "SET 3"), one can see that the first two significant models contain only the effects B, A, and AB at all three significance levels α used as input. Model III (Set 3), however, leads to a significant model (at $\alpha = 0.05$ in cumulative dropping and $\alpha = 0.01$ in single dropping) containing all effects except ABC. Therefore, the user would conclude that either Model 1 or Model II was the right one to use since both led to the same "most probable significant model."

Although the problems of estimation are not discussed in the present report, it is interesting to see how close to their true values the constants are estimated in the significant model which reads as follows for both Models I and II (the printout of the regression coefficients of Step 5 for "Set 1" and "Set 2" is not reproduced):

$$\hat{Y} = 13.2 + 2.7x_1 + 10.8x_2 - 1.9x_3 + 7.8x_4 + 5.6x_6 + 3.4x_7 - 20.0x_{11}.$$

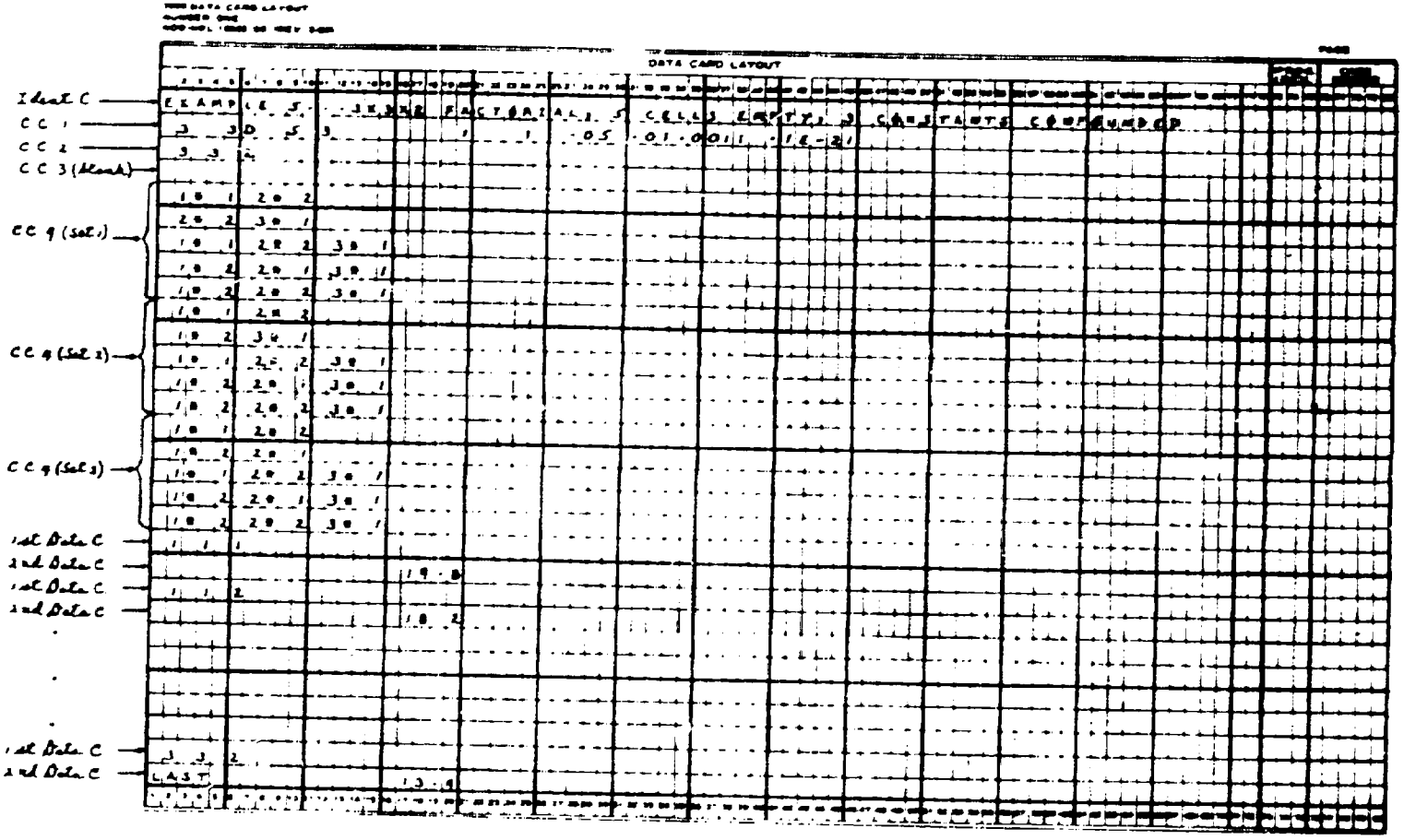
That is, one has:

$$\hat{m} = 13.2$$

$$\hat{a}_1 = 2.7 \qquad \hat{b}_1 = -1.9$$
$$\hat{a}_2 = 10.8 \qquad \hat{b}_2 = 7.8$$

$$\hat{ab}_{11} = 5.6$$
$$\hat{ab}_{21} = -20.0 \qquad \hat{ab}_{22} = 3.4$$

None of these estimates is significantly different from the true values (which were listed earlier) when testing at the 0.05 significance level.


DATA CARD LAYOUT
Ident C
CC 1
CC 2
CC 3 (blank)
CC 4 (set 1)
CC 4 (set 2)
CC 4 (set 3)
1st Data C
2nd Data C
1st Data C
2nd Data C
1st Data C
2nd Data C
EXAMPLE 5
LAST

EXAMPLE 5...3x3x2 FACTORIAL, 5 CELLS EMPTY, 3 CONSTANTS CONFOUNDED
/FEMP/Y1/CUMUL /SET 01

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	00001X00002X00003		0.38548652	.77884615	001	.90353382	.77884615	001	.86200000	005	.99689336
002	00002X00003		0.59544991	.21436813	001	.57611037	.49660714	002	.86200000	005	.99633197
003	00001X00003		0.61186494	.75789286	002	.72360789	.62375000	004	.86200000	005	.99617745
004	00003		0.57185553	1.1400000	001	.84338747	.72700000	005	.86200000	005	.99509496
005	00001X00002	*	0.00020190	121.74250	003	53.489342	46.107813	008	.86200000	005	.99427329
006	00001		0.00010697	105.97899	002	67.380566	58.082048	010	.86200000	005	.73181766
007	00002		0.00001872	401.10948	002	133.70451	115.25329	012	.86200000	005	.57823858

/FEMP/Y1/CUMUL /SET 02

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	00001X00002X00003		0.38548652	.77884615	001	.90353382	.77884615	001	.86200000	005	.99689336
002	00002X00003		0.70038839	.25281177	002	.49670139	.42815657	003	.86200000	005	.99633197
003	00001X00003		0.61186494	1.2105903	001	.72360789	.62375000	004	.86200000	005	.99596751
004	00003		0.57185553	1.1400000	001	.84338747	.72700000	005	.86200000	005	.99509496
005	00001X00002	*	0.00020190	121.74250	003	53.489342	46.107813	008	.86200000	005	.99427329
006	00001		0.00010697	105.97899	002	67.380566	58.082048	010	.86200000	005	.73181766
007	00002		0.00001872	401.10948	002	133.70451	115.25329	012	.86200000	005	.57823858

/FEMP/Y1/CUMUL /SET 03

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	00001X00002X00003		0.38548652	.77884615	001	.90353382	.77884615	001	.86200000	005	.99689336
002	00001X00002	*	0.02599039	9.4552292	002	7.6138068	6.5631015	003	.86200000	005	.99633197
003	00001X00003	*	0.00014198	133.80990	002	66.661847	57.461823	005	.86200000	005	.98270133
004	00002X00003		0.00015925	37.579427	002	60.070926	51.781138	007	.86200000	005	.78980125
005	00003		0.00020190	6.3945355	001	53.489342	46.107813	008	.86200000	005	.73562683
006	00001		0.00010697	105.97899	002	67.380566	58.082048	010	.86200000	005	.73181766
007	00002		0.00001872	401.10948	002	133.70451	115.25329	012	.86200000	005	.57823858

Test [illegible], Example 3, Continued:

/FEMB/V1/SINGLE/SET 04

STEP	EFFECT
001	0000_1*00002*00003
002	00002*00003
003	0000_1*00003
004	00003
005	0000_1*00002
006	00001
007	00002

/FEMB/V1/SINGLE/SET 04

STEP	EFFECT
001	0000_1*0000_2*00003
002	00002*00003
003	0000_1*0000_3
004	00003
005	0000_1*00002
006	00001
007	00002

/FEMB/V1/SINGLE/SET 03

STEP	EFFECT
001	00001*00002*0000$_3$
002	0000_1*00002
003	0000_1*0000_3
004	00002*00003
005	00003
006	00001
007	00002

PRC	I(X)	DIFF MS	DIF	DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
	0.38548852	.77884615	00_1	.00353382	.77884615	00_1	.86200000	005	.99689336	
	0.43306124	.21436813	00_1	.25275057	.21436813	00_1	.84814103	006	.99633197	
	0.41781946	.75089286	00_2	.99154418	.75089286	00_2	.75760204	007	.99617749	
	0.25063210	1.1400000	00_1	1.5077148	1.1400000	00_1	.75611111	009	.99589496	
•	0.00000001	121.74250	003	153.23199	121.74250	003	.79450000	010	.99427329	
	0.05373657	105.97899	00_2	3.6019302	105.97899	00_2	28.705977	013	.73101744	
	0.00154171	401.10948	00_2	10.282565	401.10948	00_2	39.008698	015	.97823858	

PRC	I(X)	DIFF MS	DIF	DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
	0.38548852	.77884615	00_1	.00353382	.77884615	00_1	.86200000	005	.99689336	
	0.75262941	.25281177	002	.29807752	.25281177	002	.84814103	006	.99633197	
	0.22472976	1.2109303	001	1.7310585	1.2109303	001	.69930871	008	.99598751	
	0.25063210	1.1400000	00_1	1.5077148	1.1400000	00_1	.75611111	009	.99589496	
•	0.00000001	121.74250	003	153.23199	121.74250	003	.79450000	010	.99427329	
	0.05373657	105.97899	002	3.6019302	105.97899	002	28.705977	013	.73101744	
	0.00154171	401.10948	00_2	10.282565	401.10948	00_2	39.008698	015	.97823858	

PRC	I(X)	DIFF MS	DIF	DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
	0.38548852	.77884615	001	.00353382	.77884615	001	.86200000	005	.99689336	
•	0.00953371	9.4552292	00_2	11.140188	9.4552292	00_2	.84814103	006	.99633197	
•	0.00004887	133.80990	00_2	44.604994	133.80990	00_2	8.9999131	00d	.99570133	
	0.31773403	37.579427	00_2	1.2886478	37.579427	00_2	29.161911	010	.78000185	
	0.45555810	8.3949355	00_1	.20921220	8.3949355	00_1	38.564831	012	.73562403	
	0.05373657	105.97899	002	3.6019302	105.97899	002	28.705977	013	.73101744	
	0.00154171	401.10948	00_2	10.282565	401.10948	00_2	39.008698	015	.97823858	

Problem running time: 1 min. 25 sec.

3.4.6 Example 6

Among all examples exhibited in the present report, Example 6 shows the largest number of NOVACOM features in combination. Example 6 is one of analysis of covariance for an unbalanced and incomplete 3x3x3 data classification with 2 dependent variables and 3 OCIVs. The data is synthetic.

Of the 27 cells of the layout 9 were randomly selected to be empty. Factors α and γ are quantitative, and factor β is qualitative. The quantitative factor level values are unequally spaced and their values are coded as are those of the 3 OCIVs.

Two different models were used in the construction of the data for the two dependent variables: For Y_1 a model was used in which γ is a dummy factor, whereas for Y_2 a model was used where α is a dummy factor. The constants of the two models are as follows:

For Y_1:

$$m = 13$$

$$a_1 = 5 \qquad b_1 = -6$$
$$a_2 = 9 \qquad b_2 = 3$$

$$ab_{11} = 2 \qquad ab_{12} = 40$$
$$ab_{21} = -10$$

For Y_2:

$$m = 13$$

$$b_1 = 20 \qquad c_1 = 5$$
$$b_2 = 10 \qquad c_2 = 25$$

$$bc_{11} = 1 \qquad bc_{12} = 30$$
$$bc_{21} = -16 \qquad bc_{22} = 60$$

(Note. The constant ab_{22} was not needed in the model for Y_1; see further below.)

Table 3.6 shows the data layout of the values of Y_1 and Y_2, i.e., of the expected values of the response variables. Also shown are the numbers of repeated observations in the cells, $R_{\alpha\beta\gamma}$, and the values (factor levels) of the quantitative factor variables X_α and X_γ.

		C_1 $X_{c1}=1$	C_2 $X_{c2}=2$	C_3 $X_{c3}=9$
A_1 $X_{a1}=2$	B_1	$R_{111}=2$ $Y_1=14$ $Y_2=39$	$R_{112}=1$ $Y_1=14$ $Y_2=88$	$R_{113}=2$ $Y_1=14$ $Y_2=33$
	B_2	$R_{121}=0$	$R_{122}=4$ $Y_1=61$ $Y_2=108$	$R_{123}=0$
	B_3	$R_{131}=2$ $Y_1=18$ $Y_2=18$	$R_{132}=1$ $Y_1=18$ $Y_2=38$	$R_{133}=1$ $Y_1=18$ $Y_2=13$
A_2 $X_{a2}=6$	B_1	$R_{211}=3$ $Y_1=6$ $Y_2=39$	$R_{212}=0$	$R_{213}=0$
	B_2	$R_{221}=1$ $Y_1=25$ $Y_2=12$	$R_{222}=0$	$R_{223}=1$ $Y_1=25$ $Y_2=23$
	B_3	$R_{231}=1$ $Y_1=22$ $Y_2=18$	$R_{232}=3$ $Y_1=22$ $Y_2=38$	$R_{233}=1$ $Y_1=22$ $Y_2=13$
A_3 $X_{a3}=7$	B_1	$R_{311}=2$ $Y_1=7$ $Y_2=39$	$R_{312}=1$ $Y_1=7$ $Y_2=88$	$R_{313}=2$ $Y_1=7$ $Y_2=33$
	B_2	$R_{321}=0$	$R_{322}=0$	$R_{323}=0$
	B_3	$R_{331}=1$ $Y_1=13$ $Y_2=18$	$R_{332}=1$ $Y_1=13$ $Y_2=38$	$R_{333}=0$

Table 3.6

Data Layout Example 6

Considering, for the moment, all three factors as being qualitative, Table 3.7 shows the fitting of ABC-constants using the method explained in Appendix A. Only one ABC-constant may be fitted; abc_{111} was selected from the two possible constants (the other one being abc_{112}).

		C_1	C_2	C_3
A_1	B_1	(X)	X	X
	B_2		X	
	B_3	X	X	X
A_2	B_1	X		
	B_2	X		X
	B_3	X	X	X
A_3	B_1	X	X	X
	B_2			
	B_3	X	X	

Table 3.7

Fitting of ABC-Constants in Example 6

Since factors A and C are, in reality, quantitative factors, the fitted constant abc_{111} must be interpreted accordingly. One can easily see that abc_{111} can represent the interaction between the component $a_{1\,1_a} \times c_{1\,1_a}$ and factor B. Therefore, abc_{111} is equivalent to $X_a b_1 X_c$ with 1 degree of freedom. In the terms of the NOVACOM notation this is the DIV 1.1 x 2*1 x 3.1. Since this is the only DIV representing the interaction ABC (containing both qualitative and quantitative factors), this DIV is also a PFFE.

Tables 3.8 a-c show the fitting of two-factor interaction constants. Again, all three factors are considered to be qualitative for the moment. As can be seen, 3 of the 4 AB-constants can be fitted, and all 4 AC- and all 4 BC-constants can be fitted. Since all 6 main effect constants can be fitted, one has 6 + 11 + 1=18 constants which appear as if they can be fitted. However, there are only 27 - 9 - 1 = 17 degrees of freedom "between cells." Consequently, there must be one identity in the data. Looking at Table 3.7, one notes at once that eliminating all observations from cell A_1B_2 also eliminates all observations from cell B_2C_2. Therefore, the identity is:

$$A_1B_2 \equiv B_2C_2$$

	B_1	B_2	B_3
A_1	(X)	(X)^I	
A_2	(X)		
A_3			

	C_1	C_2	C_3
A_1	(X)	(X)	
A_2	(X)	(X)	
A_3			

	C_1	C_2	C_3
B_1	(X)	(X)	
B_2	(X)	(X)^I	
B_3			

Tables 3.8 a-c

Fitting of the Two-Factor Interaction Constants in Example 6

One can fit either ab_{12} or bc_{22}, but not both at the same time.

In order not to add unnecessarily to the amount of printout to be reproduced, it was decided to have only one CC 4 set and to fit bc_{22}. (The fitting of bc_{22} leaves the interaction AB, which is represented by 2 constants, to be the only other PFFE in addition to ABC.) Since C is a dummy factor for Y_1, the fitting of bc_{22} (instead of ab_{12}) should lead to a false significant model containing factorial effects which involve factor C. For Y_2, A is a dummy factor, and the fitting of bc_{22} (instead of ab_{12}) should lead to a realistic significant model. These assumptions are verified by the FEMO rankings, see further below.

The interpretation of the 17 fitted constants for the real situation of factors a and c being quantitative is not difficult. The relations for all 17 constants fitted are as follows:

Fit X_a and X_a^2 instead of a_1 and a_2.

Fit X_c and X_c^2 instead of c_1 and c_2.

Fit $X_a b_1$ and $X_a^2 b_1$ instead of ab_{11} and ab_{21}.

Fit $X_a X_c$, $X_a X_c^2$, $X_a^2 X_c$, $X_a^2 X_c^2$ instead of ac_{11}, ac_{12}, ac_{21}, ac_{22}.

Fit $b_1 X_c$ and $b_1 X_c^2$ instead of bc_{11} and bc_{12}.

Fit $b_2 X_c$ and $b_2 X_c^2$ instead of bc_{21} and bc_{22}.

Fit $X_a b_1 X_c$ instead of abc_{111}.

The 17 DIVs were "hand"-generated via CC No. 4; see the reproduced input sheet. A third order model in the three OCIVs was automatically generated from which 9 GCIVs were deleted via CC No. 5, leading to a total of 10 covariates (CIVs) in this analysis of covariance example.

The residual terms, e, in the observed values, y = Y+e, of the two dependent variables were taken from a table of normal random deviates with σ=1. The 30 values each of y_1 and y_2 and the 30 values each of the three OCIVs are given on the reproduced input sheet.

The significant CIVs, if any, to be kept in the model for the FEMO ranking are determined by the choice of KALPHA = 1 and ALPHA(1) = 0.05. ALPHA(2) = 0.01 may be considered as the principal significance level in this example. See also the discussion in Section 3.1.3. All rankings in the present example are performed under restricted admissibility rules (CAD = 0 in column 38 of CC No. 1).

The printout exhibited for Example 6 consists of the identification of the IVs, all FCAs, and the ranking information for the ANVA of OCIV 1. Following are the notes referring to the printout.

Notes on printout Example 6.

Note 6.1. The coded factor levels (values of the quantitative factor variables) are printed. For example, $X'_{a3} = \frac{X_{a3} - \bar{X}_a}{R} = \frac{7-5}{5} = 0.4$.

Note 6.2. The identification of the 17 IVs and the 14 effects is reproduced since an example of the identification for which the number of IVs is different from the number of effects was not shown previously.

(Also, the numbering of the effects is necessary information if the reader wants to follow the ranking process in the ANVA for OCIV No. 1; see Note 6.11 further below.) Note that effects Nos. 6 and 7 represent the PFFE $\mathcal{AB}$ with 2 degrees of freedom and that effect No. 14 represents the PFFE $\mathcal{ABC}$ with 1 degree of freedom.

Note 6.3. For y_1, only OCIV No. 1 is significant at the 0.05 level. Note that due to the fact of having fitted 10 CIVs there are only 2 degrees of freedom for error: DF(2) = 2.

With OCIV No. 1 being significant at ALPHA(1) = 0.05, this OCIV will be kept in FEMO and, therefore, the program will perform the ANVAs for y_1 and OCIV No. 1.

Note 6.4. COMO, single dropping, yields a significant model (at $\alpha = 0.05$) containing 4 CIVs. However, with only 2 degrees of freedom for error in COMO, cumulative dropping, it does not make sense to try to close the gap between the two models. (In practice, one would not fit such a large covariate-model as was done here for demonstration purposes.)

Note 6.5. The significant model for y_1 resulting from FEMO, cumulative dropping, contains four factorial effects involving the dummy factor $\mathcal{C}$, as was predicted. DF(2) equals 11 after one degree of freedom for OCIV No. 1 was subtracted from the degrees of freedom "within cells."

Note 6.6. The single dropping procedure of FEMO for y_1 results in the same significant model as was obtained with the cumulative procedure.

Note 6.7. For the second dependent variable, y_2, COMO, cumulative dropping, does not show any significant CIVs. Therefore, no ANVA will be performed for y_2 or any other OCIV than No. 1.

Note 6.8. The single dropping procedure of COMO for y_2 does show significant CIVs, however, again because of only 2 degrees of freedom for error in the cumulative dropping procedure, closing the gap between the two models is not worthwhile trying.

Note 6.9. FEMO, cumulative dropping, yields a significant model for y_2 which contains effects involving factors $\mathcal{B}$ and $\mathcal{C}$ only, as was predicted. That is, there are no factorial effects in the significant model involving factor $\mathcal{A}$, which is a dummy factor for y_2. The significant model is reached rather abruptly at Step 10: the $^{++}$-procedure had to be applied in order to continue the ranking. The ranking order within the significant model shows factor $\mathcal{C}$ to be by far the more important of the two factors.

Note 6.10. The single dropping procedure of FEMO results in a significant model for y_2 which contains, at the 0.05 level, also the two

degrees of freedom representing the main effect of the dummy factor $\mathcal{D}$. Indeed, dividing DIFF MS = 4.0072434 of Step 8 in FEMO, cumulative dropping, by MS(2) = .68833333 yields F = 5.821 which, with 1 and 12 degrees of freedom, is marginally significant at the 0.05 level. However, because of this marginal significance (which actually is random, as is known from the construction of the data!) and because ALPHA(2) = 0.01 was decided upon in advance to be used as the principal significance level, one may say that both dropping procedures show the same significant model (containing effects involving $\mathcal{B}$ and $\mathcal{C}$ only).

Note 6.11. This is the printout of the ranking information for the ANVA of OCIV No. 1. (This and the FCA is the only information ever given for any ANVA.)

Note 6.12. In the last step of the FEMO-type ranking for OCIV No. 1, the **-procedure had to be applied. (For practical purposes, this has no influence upon the ranking here since effect No. 3 is the only one left not yet ranked and, thereby, is the most important effect by definition.) The last three I(X)-values are those of "Step 14", "Step 14*", and "Step 14**."

Note 6.13. In the identification of the ANVA printout the cardinal number of the OCIV or the symbol of the dependent variable, Y, is given. Since the present ANVA is that for OCIV No. 1, the identification "$\mathcal{Y}$01" is printed.

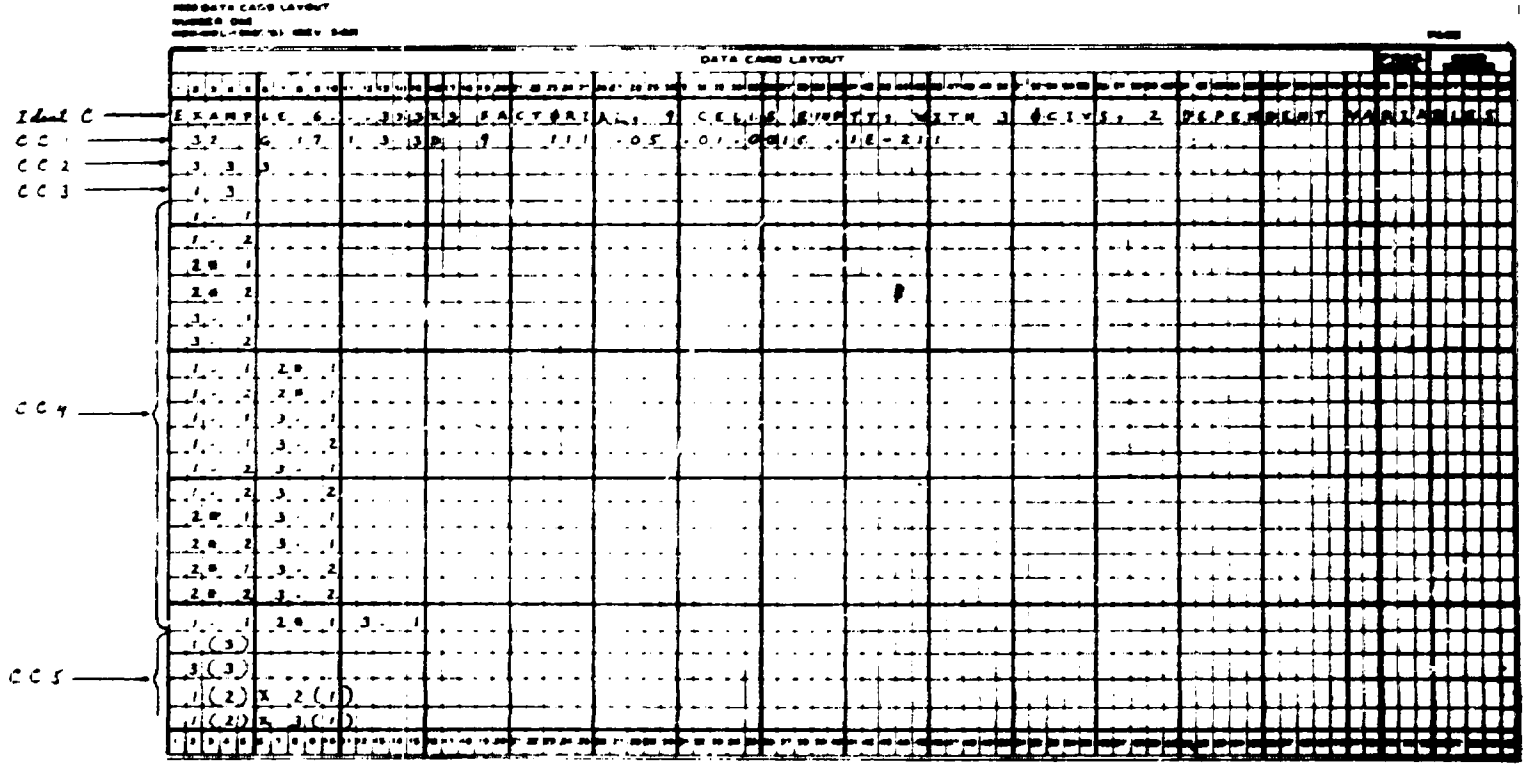
Note 6.14. The FCA, ANVA (cumulative dropping) for OCIV No. 1 shows that the factors and their interactions, in the present example, had significant effects upon this concomitant independent variable. Except for the interaction $\mathcal{AB}$, all factorial effects contained in the significant model for y_1 in the analysis of covariance (see Note 6.5) are also contained in the significant model (at α = 0.01) for OCIV No. 1. This happens because the numerical values of OCIV No. 1 were constructed such that they are highly correlated with the values of y_1.

Note 6.15. This is the FCA, ANVA, for y_1. (The preceding ranking information is not exhibited here.) In other words, the FCA shown is that which would have been obtained for y_1 if no OCIVs had been included in the FEMO ranking. (See the degrees of freedom for error: DF(2) = 12 which is the number of degrees of freedom for "within cells.")

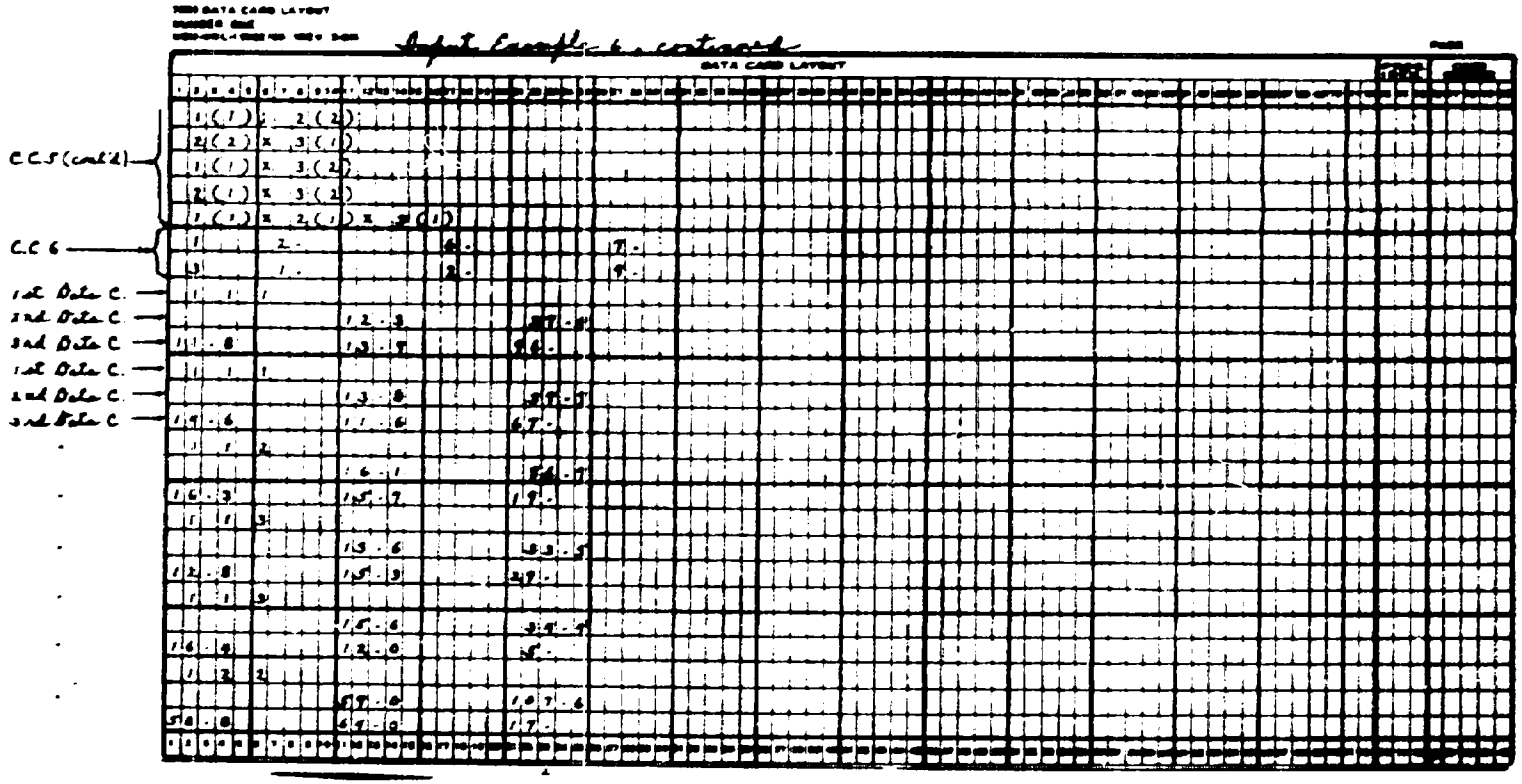
The ranking order within the significant model for y_1 alone is slightly different from that obtained in the analysis of covariance (see Note 6.5), but both significant models contain the same set of effects. It is obvious that the significance of the factorial effects is much higher when OCIV No. 1 is excluded from the model. That is, the present example shows how the use of covariates can cause a decrease in the power of the F-test although the residual variance is considerably reduced (from MS(2)= 1.0[illegible]00 with 12 degrees of freedom to MS(2) = .11673482 with 11 degrees

of freedom in the present case). Without having the ANVAs available the analyst would not know whether the factors had effects upon the covariate(s) nor whether the sensitivity of the analysis was decreased by performing an analysis of covariance ranking rather than an analysis of variance ranking..

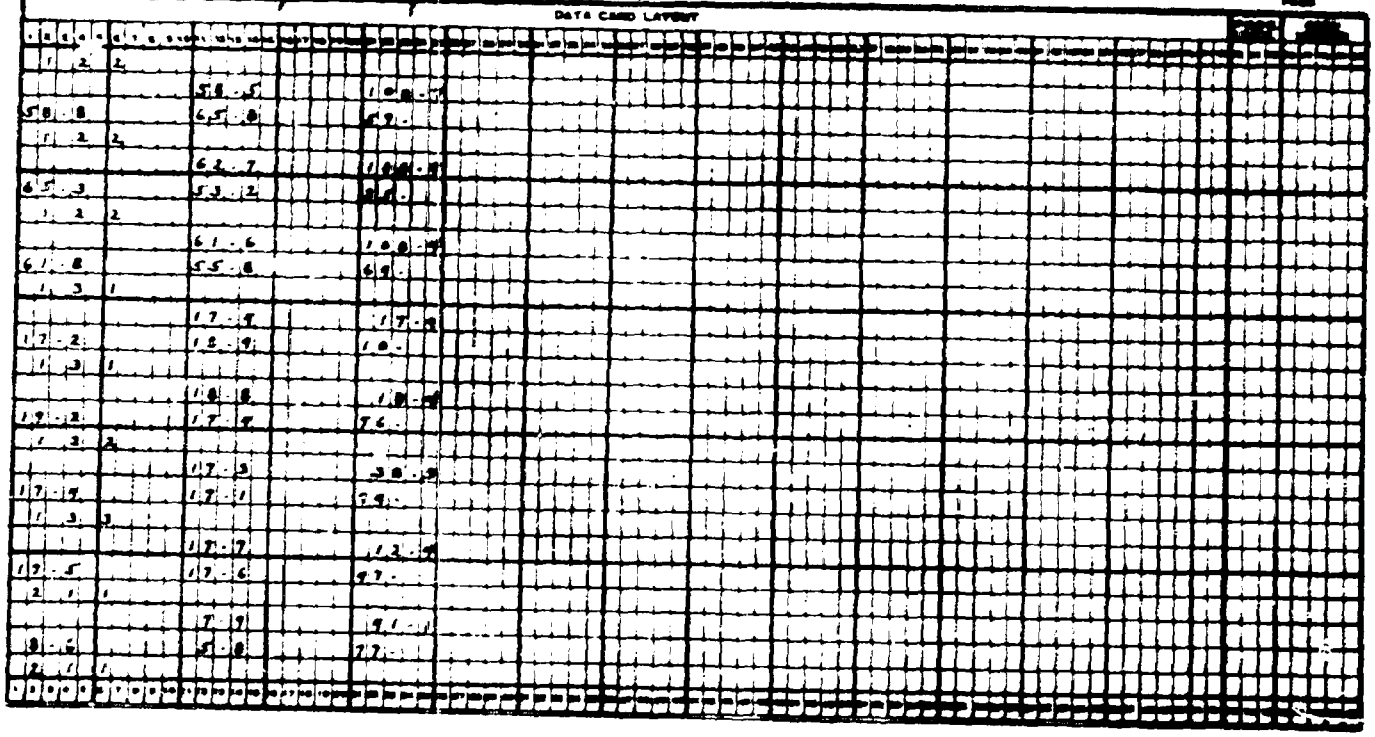
Note 6.16. The "Problem Running Time" of 4 minutes and 11 seconds is that for both dependent variables and includes the time for the 2 ANVAs.


DATA CARD LAYOUT
Ident C
CC 1
CC 2
CC 3
CC 4
CC 5

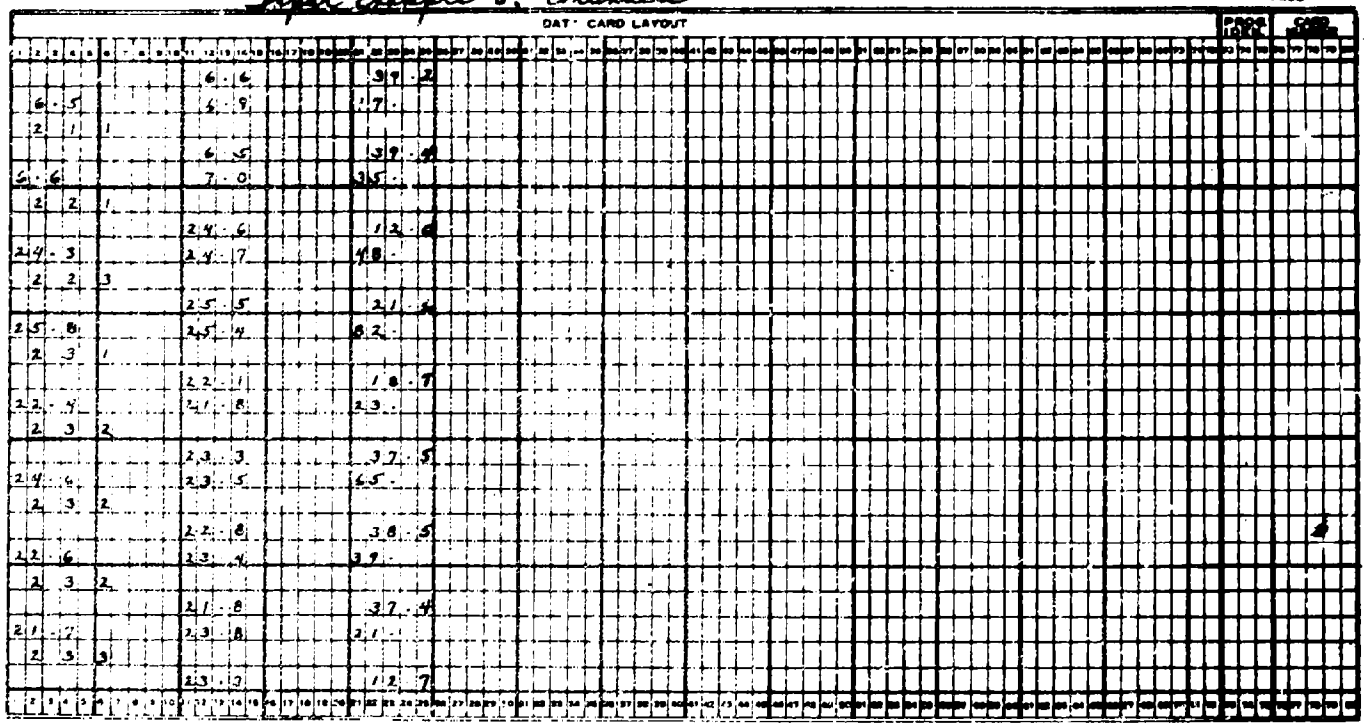
Input Example 4 continued



Input Example continued

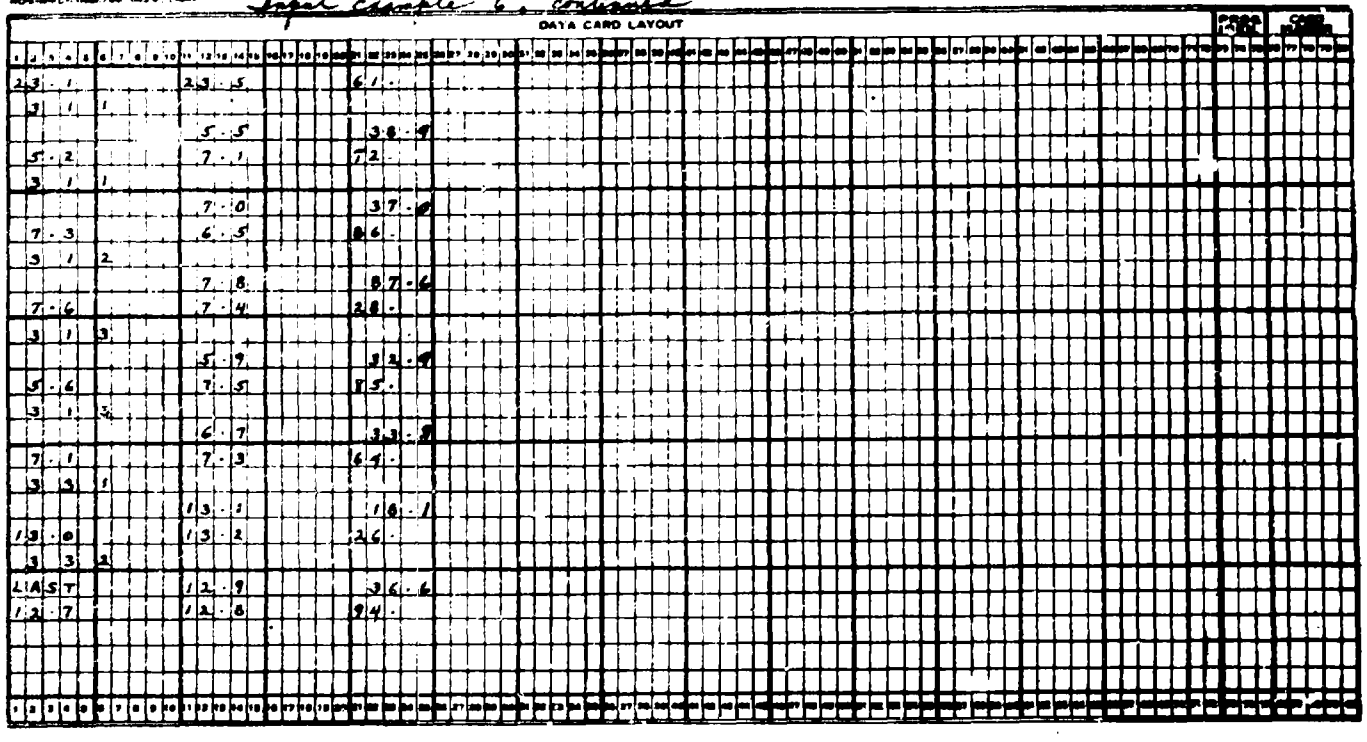


Input Example 6, continued



7090 DATA CARD LAYOUT

Input Example 6, continued



1st Data C.
2nd Data C.
3rd Data C.

```
NZVACAM ... NON-ORTHOGONAL VARIANCE AND COVARIANCE ANALYSIS BY MULTIPLE REGRESSION ... PROGRAM VERSION OF JULY 15,1967

 I  VARDV  U   SUB  NCC4 NSCC4  TP  P GDC NCC5 CODE     C     DRAP  NALPHA ALPHA(1) ALPHA(2) ALPHA(3)  CAD     TPLT2     IRC4 ANA
 01    2   -0    ,   DIF    1      4 03  3    9    -  001.00     1      1    .0500    .0100    .0010     0   .1000E-02     1   1

QUANTITATIVE FACTOR LEVEL VALUES
FACTOR NO

     1       .20000000000E+01 0.40000 00000E+ 1 .60000000000E+01

     2 DEL  -.60 00000000E+00 0.20000 0000E+ 0  .40000000000E+00   Note 6.1
AVERAGE  0.60000000000000E+01  RANGE  0.80 00000000000E+01

     3       0.10000000000E+01 0.20000 0000E+01 .30000000000E+01

     4 DEL  -.37500000000E+00 -.25000 0000E- 1  0.62500000000E+00
AVERAGE  0.40000000000000E+01  RANGE  0.80000000000000E+01
  1  1 -1  1  1  2 -1  2  2  3  2  1  2  3  2  2  3  4 -3  1  3 -3 -3  2 -4  6 -1  1  2  1 -4  7 -1  2  2  1  9  8 -1  1 -1  1  5
  9 -1  1 -3  2  9 10 -1  2 -3  1  9 11 -1  2 -3  2  6 12  2  1 -3  1  6 12  2  2 -3  1  6 13  2  1 -3  2  6 14  2  2 -3  2 -7 14
 -1  1  2  1 -3  1  0  1  0  2  0  3  0  1  1  0  1  2  0  1  3  0  2  2  0  2  3  0  3  3  0 -1  1  1  0 -1  1  2  0 -1  1  3  0
 -1  2  2  0 -1  2  0  0 -1  3  3  .  2  2  2  0 -2  2  3  0 -2  3  3  0 -3  3  3
EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 DEPS, 2 DEPENDENT VARIABLES
/    /  /      /SET  1

IDENTIFICATION OF INDEPENDENT VARIABLES (IV-S)

DIV-S...

        EFFECT NUMBER  1                Note 6.2

        IV   1      1. 1

        EFFECT NUMBER  2

        IV   2      1. 2

        EFFECT NUMBER  3

        IV   3      2. 1
        IV   4      2. 2

        EFFECT NUMBER  4

        IV   5      3. 1

        EFFECT NUMBER  5

        IV   6      3. 2

        EFFECT NUMBER  6

        IV   7      1. 1X 2. 1

        EFFECT NUMBER  7

        IV   8      1. 2X 2. 1

        EFFECT NUMBER  8

        IV   9      1. 1X 3. 1
```

"Identification of independent variables", (example 6), continued:

```
     EFFECT NUMBER  9

     IV  10      I, 1X 3, 2

     EFFECT NUMBER 10

     IV  11      I, 2X 3, 1

     EFFECT NUMBER 11

     IV  12      I, 2X 3, 2

     EFFECT NUMBER 12

     IV  13      Z* 1X 3, 1
     IV  14      Z* 2X 3, 1

     EFFECT NUMBER 13

     IV  15      Z* 1X 3, 2
     IV  16      Z* 2X 3, 2

     EFFECT NUMBER 14

     IV  17      I, 1X 2* 1X 3, 1

CIV=S...

     IV  18      I( 1)
     IV  19      Z( 1)
     IV  20      S( 1)
     IV  21      I( 2)
     IV  22      I( 1)X 2( 1)
     IV  23      I( 1)X 3( 1)
     IV  24      Z( 2)
     IV  25      Z( 1)X 3( 1)
     IV  26      S( 2)
     IV  27      Z( 3)
```

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 #CIVS, 2 DEPENDENT VARIABLES
/COMB/Y1/CUMUL /SET 1

FINAL COMPREHENSIVE COMB-CUMULATIVE DROPPING

Note 6.3

STEP	CONC VAR
1	1(1)X 3(1)
2	3(2)
3	1(2)
4	2(1)X 3(1)
5	2(3)
6	3(1)
7	2(2)
8	1(1)X 2(1)
9	2(1)
10	1(1)

PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFFI DET
	0.92202384	.00102484	1	.01223502	.00102684	1	.08392643	2	.99997984
	0.98400131	.00170225	1	.01625881	.00136454	2	.08392643	2	.99997971
	0.98893660	.00604034	1	.03482982	.00292314	3	.08392643	2	.99997878
	0.96446846	.03021995	1	.11614152	.00974734	4	.08392643	2	.99997878
	0.97189716	.01390274	1	.12604400	.01057842	5	.08392643	2	.99997515
	0.93156077	.06329042	1	.23072296	.01936376	6	.08392643	2	.99997348
	0.68782966	.30915803	1	.72400237	.06076294	7	.08392643	2	.99996888
	0.63882913	.15356850	1	.86222695	.07236363	8	.08392643	2	.99992874
	0.48761855	.53732112	1	1.4777892	.12402558	9	.08392643	2	.99991029
*	0.04080999	18.942584	1	23.900472	2.00588i4	10	.08392643	2	.99984874

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 #CIVS, 2 DEPENDENT VARIABLES
/COMB/Y1/SINGLE/SET 1

FINAL COMPREHENSIVE COMB-SINGLE DROPPING

Note 6.4

STEP	CONC VAR
1	1(1)X 3(1)
2	3(2)
3	1(2)
4	2(1)X 3(1)
5	2(3)
6	3(1)
7	2(2)
8	1(1)X 2(1)
9	2(1)
10	1(1)

PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFFI DET
	0.92202384	.00102484	1	.01223502	.00102684	1	.08392643	2	.99997984
	0.87302132	.00170925	1	.03023892	.00170225	1	.05629324	3	.99997971
	0.72576688	.00604034	1	.14164077	.00604034	1	.04264549	4	.99997891
	0.39744724	.03021995	1	.85549633	.03021995	1	.03532446	5	.99997878
	0.54883350	.01390274	1	.40328544	.01390274	1	.03447371	6	.99997515
	0.19950543	.06329042	1	2.0069897	.06329042	1	.03153500	7	.99997348
*	0.01840140	.30915803	1	8.7075915	.30915803	1	.03550443	8	.99996888
	0.16124925	.15356850	1	2.3299593	.15356850	1	.0699i038	9	.99992874
	0.02299721	.53732112	1	7.1953669	.53732112	1	.07467619	10	.99991029
*	0.00000006	18.942584	1	162.27020	18.942584	1	.11473482	11	.99984874

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 DEPVS, 2 DEPENDENT VARIABLES
/FEMB/Y1/CUMUL /SET 01

FINAL COMPREHENSIVE FEMB-CUMULATIVE DROPPING Note 6.5

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X00002X03.01		0.41882336	.08235650	001	.70550068	.08235650	001	.11673482	011	.99884574
002	01.02X03.02		0.62596115	.03180939	00_1	.48899472	.05708295	00_2	.11673482	011	.99889589
003	01.01X03.02		0.00316950	.00175535	00_1	.33101017	.03864041	003	.11673482	011	.99883203
004	01.02X03.01		0.54344155	.26299398	00_1	.81148712	.09472881	004	.11673482	011	.99883182
005	01.01X03.01		0.61968774	.04344021	00_1	.72361515	.08447109	005	.11673482	011	.99880022
006	01.02X00002	*	0.00477380	3.9027155	00_1	6.1750693	.72084515	006	.11673482	011	.99979900
007	01.01X00002		0.00308687	1.1188474	00_1	6.6621304	.77770251	007	.11673482	011	.99932617
008	01.02		0.00347593	.36592098	00_1	6.2211934	.72622991	008	.11673482	011	.99919174
009	00002X03.02		0.00374726	.48993715	002	5.8163566	.67897136	010	.11673482	011	.99916780
010	00002X03.01		0.00463534	.34128810	00_2	5.3342336	.62269082	012	.11673482	011	.99903009
011	03.02		0.00573095	.10785908	00_1	4.9949823	.58308857	013	.11673482	011	.99894809
012	03.01		0.00729144	.03552770	00_1	4.69999a8	.54397690	014	.11673482	011	.99893513
013	01.01		0.00830057	.18698738	00_1	4.4560619	.52017760	015	.11673482	011	.99893087
014	00002		0.00937249	.31477909	00_2	4.2490512	.49601224	017	.11673482	011	.99890648

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 BCIVS, 2 DEPENDENT VARIABLES
/FEMB/Y1/SINGLE/SET A1

FINAL COMPREHENSIVE FEMB-SINGLE DROPPING

Note - 6

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X01002X03.01		0.41882336	.08235850	001	.70550068	.08235850	001	.11673482	011	.99884574
002	01.02X03.02		0.60676579	.03180939	00_1	.27934841	.03180939	00_1	.11386996	012	.99889585
003	01.01X03.02		0.90030177	.00175535	00_1	.01632009	.00175535	00_1	.10755761	013	.99893203
004	01.02X03.01		0.12716256	.26299398	00_1	2.4299918	.26299398	00_1	.10000051	014	.99886182
005	01.01X03.01		0.94075061	.04344021	00_1	.39182429	.04344021	00_1	.11086655	015	.99880022
006	01.02X00002	*	0.00001688	3.9027955	00_1	36.592890	3.9027155	00_1	.10669241	016	.99179900
007	01.01X00002		0.08307251	1.1188474	00_1	3.3909580	1.1188474	00_1	.32995023	017	.99932617
008	01.02		0.33556570	.36592098	00_1	.97897083	.36592098	00_1	.37377785	018	.99919176
009	00002X03.02		0.29253775	.48993715	00_2	1.3122227	.48993715	00_2	.37336433	019	.99914780
010	00002X03.01		0.42649347	.34128810	00_2	.88769267	.34128810	00_2	.38446451	021	.99909009
011	03.02		0.59964531	.10785908	00_1	.28330899	.10785908	00_1	.38071186	023	.99896809
012	03.01		0.75912583	.03552770	00_1	.09619162	.03552770	00_1	.36934300	024	.99895513
013	01.01		0.47535571	.18698738	00_1	.52525964	.18698738	00_1	.35599038	025	.99893087
014	00002		0.41860353	.31477209	00_2	.90064053	.31477209	00_2	.34949027	026	.99890840

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 BCIVS, 2 DEPENDENT VARIABLES
/CBMB/V2/CUMUL /SET 1

FINAL COMPREHENSIVE COMB-CUMULATIVE DROPPING

Note 6.7

STEP	CONC VAR
1	1(2)
2	1(1)X 2(1)
3	2(3)
4	2(1)X 3(1)
5	1(1)X 3(1)
6	1(1)
7	3(2)
8	3(1)
9	2(2)
10	2(1)

PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
	0.99398416	.00002788	1	.00007238	.00002788	1	.38916938	2	.99997175
	0.87612952	.11089722	1	.14399522	.05546255	2	.38916938	2	.99997175
	0.41083157	1.7106471	1	1.9764252	.60719073	3	.38916938	2	.99998769
	0.42420482	.60201981	1	1.5730690	.60589800	4	.38916938	2	.99998496
	0.48876125	.07890876	1	1.2994287	.50050015	5	.38916938	2	.99988268
	0.52895435	.19831467	1	1.1686700	.45013590	6	.38916938	2	.99987999
	0.42045237	1.8665444	1	1.8940079	.69247998	7	.38916938	2	.99987272
	0.45492998	.10444996	1	1.5141543	.58397623	8	.38916938	2	.99988428
	0.38716941	2.0293592	1	1.9331088	.74457434	9	.38916938	2	.99980045
	0.38707034	.78849222	1	1.9445111	.74896613	10	.38916938	2	.99972603

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 BCIVS, 2 DEPENDENT VARIABLES
/CBMB/V2/SINGLE/SET 1

FINAL COMPREHENSIVE COMB-SINGLE DROPPING

Note 6.8

STEP	CONC VAR
1	1(2)
2	1(1)X 2(1)
3	2(3)
4	2(1)X 3(1)
5	1(1)X 3(1)
6	1(1)
7	3(2)
8	3(1)
9	2(2)
10	2(1)

PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
	0.99398416	.00002788	1	.00007238	.00002788	1	.38916938	2	.99997175
	0.55799415	.11089722	1	.43186167	.11089722	1	.29478868	3	.99997175
	0.04948922	1.7106471	1	7.7645170	1.7106471	1	.22031597	4	.99998769
	0.33041056	.60201981	1	1.1613435	.60201981	1	.51838219	5	.99998496
	0.71350602	.07890876	1	.14823507	.07890876	1	.53232179	6	.99988268
	0.93565656	.19831467	1	.42415851	.19831467	1	.46794850	7	.99987999
	0.07177168	1.8665444	1	4.3018416	1.8665444	1	.43389427	8	.99987272
	0.66457355	.10444996	1	.17811516	.10444996	1	.59307762	9	.99988428
	0.08229085	2.0293592	1	3.7286469	2.0293592	1	.34421486	10	.99980045
	0.30434747	.78849222	1	1.1608453	.78849222	1	.67922798	11	.99972603

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 SCIVS, 2 DEPENDENT VARIABLES
/FEMS/V2/CUMUL /SET N1

FINAL COMPREHENSIVE FEMS-CUMULATIVE DROPPING Note 6.9

STEP	EFFECT	PRC	F(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X00002X03.01		0.11450589	1.9938462	001	2.8986288	1.9938462	001	.68833333	012	.99969712
002	01.02X03.02		0.29365646	.12794748	001	1.5412545	1.0608968	002	.68833333	012	.9996240[illegible]
003	01.01X03.02		0.41500434	.00012960	001	1.0275697	.70730774	003	.68833333	012	.99961932
004	01.02X03.01		0.55391530	.05171966	001	.78945868	.54341072	004	.68833333	012	.99961931
005	01. [illegible] 03.01		0.67433390	.01698719	001	.63650269	.43812602	005	.68833333	012	.99961742
006	01.02X00002		0.64915245	.73765149	001	.70897059	.48801359	006	.68833333	012	.99961680
007	01.01X00002		0.69485828	.56475888	001	.72490635	.49897721	007	.68833333	012	.99958975
008	01.02		0.38357771	4.0072434	001	1.3820087	.93751048	008	.68833333	012	.99956905
009	01.01		0.35942122	.14178186	001	1.2335489	.84908730	009	.68833333	012	.99942211
010	00002X03.02	**	0.00035322	2002.1981	002	11.454513	2002.1981	002	174.79557	023	.99941691
011	00002X03.01		0.29113013	227.67936	002	1.3025469	227.67936	002	174.79557	023	.85258334
012	00002		0.00000679	4619.6362	002	13.065671	2423.6578	004	174.79557	023	.83988620
013	03.02		0.00000002	11801.310	001	24.595522	4299.1882	005	174.79557	023	.49789971
014	03.01		0.00000002	1795.4274	001	22.170859	3879.2281	006	174.79557	023	.06436818

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 OCIVS, 2 DEPENDENT VARIABLES
/FEMB/Y2/SINGLE/SET 01

FINAL COMPREHENSIVE FEMB-SINGLE DROPPING

Note 6.10

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X00002X03.01		0.11450589	1.9938462	00_1	2.8966288	1.9938462	00_1	.68833333	012	.99969712
002	01.02X03.02		0.69367116	.12794748	00_1	.16221398	.12794748	00_1	.78875740	013	.99962401
003	01.01X03.02		0.98963898	.00012960	00_1	.00017476	.00012960	00_1	.74155669	014	.99961932
004	01.02X03.01		0.78830463	.05171966	00_1	.07472555	.05171966	00_1	.69212822	015	.99961931
005	01.01X03.01		0.87379904	.01698719	00_1	.02604987	.01698719	00_1	.65210268	016	.99961742
006	01.02X00002		0.28867491	.73745149	00_1	1.1996095	.73745149	00_1	.61474295	017	.99961680
007	01.01X00002		0.35310842	.56475888	00_1	.90861510	.56475888	00_1	.62156009	018	.99950975
008	01.02	•	0.01975760	4.0072434	00_1	6.4782318	4.0072434	00_1	.61857055	019	.99954905
009	01.01		0.67605098	.14170186	00_1	.17982374	.14170186	00_1	.78800419	020	.99942211
010	00002X03.02	•	0.00000000	2002.1981	00_2	2644.1157	2002.1981	00_2	.75722789	021	.99941691
011	00002X03.01		0.29115013	227.67936	00_2	1.3025465	227.67936	00_2	174.79557	023	.85258334
012	00002		0.00000083	4619.6362	00_2	25.804236	4619.6362	00_2	179.02627	025	.83588620
013	03.02		0.00004940	11801.310	001	23.232739	11801.310	001	507.96034	027	.49709971
014	03.01		0.17610979	1755.4274	00_1	1.9263014	1755.4274	00_1	911.29426	028	.06436018

```
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../ 14/
ARG1=   6.00  ARG2=   0.50  X= .98513927E-00  I(X)= 0.67802961704966E+00
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  7/ 11/ 13/
ARG1=   6.00  ARG2=   1.00  X= .60783219E+00  I(X)= 0.50431532912773E-01
ARG1=   6.00  ARG2=   1.00  X= .97332374E+00  I(X)= 0.85024462704839E+00
ARG1=   6.00  ARG2=   1.50  X= .71303491E-01  I(X)= 0.37344016945440E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  7/  9/ 10/ 13/
ARG1=   6.00  ARG2=   1.50  X= .45039912E-00  I(X)= 0.19163681723357E-01
ARG1=   6.00  ARG2=   1.50  X= .97225896E+00  I(X)= 0.95013973149029E+00
ARG1=   6.00  ARG2=   1.50  X= .96290965E+00  I(X)= 0.92505090284870E+00
ARG1=   6.00  ARG2=   2.00  X= .64346843E-01  I(X)= 0.46948564734789E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  7/ 10/ 13/
ARG1=   6.00  ARG2=   2.00  X= .44896388E-00  I(X)= 0.34887960338229E-01
ARG1=   6.00  ARG2=   2.00  X= .96128412E+00  I(X)= 0.97235621864268E+00
ARG1=   6.00  ARG2=   2.50  X= .62423152E-01  I(X)= 0.79889091608720E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  7/  8/ 13/
ARG1=   6.00  ARG2=   2.50  X= .62214095E-00  I(X)= 0.42411364088867E-01
ARG1=   6.00  ARG2=   2.50  X= .95974190E+00  I(X)= 0.99010188648857E+00
ARG1=   6.00  ARG2=   3.00  X= .90148702E-01  I(X)= 0.97727086484554E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  7/ 13/
ARG1=   6.00  ARG2=   3.00  X= .41175800E-00  I(X)= 0.57484057807803E-01
ARG1=   6.00  ARG2=   3.50  X= .97424534E-01  I(X)= 0.15759181160035E-05
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  6/ 13/
ARG1=   6.00  ARG2=   3.50  X= .30639479E-00  I(X)= 0.19325504997123E-01
ARG1=   6.00  ARG2=   4.00  X= .97418566E-01  I(X)= 0.25876805054118E-05
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  2/ 13/
ARG1=   6.00  ARG2=   4.00  X= .18650642E-00  I(X)= 0.21012906331429E-02
ARG1=   6.00  ARG2=   4.50  X= .56506181E-01  I(X)= 0.37031816582100E-05
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  1/ 13/
ARG1=   6.00  ARG2=   4.50  X= .14229362E-00  I(X)= 0.71373200758102E-03
ARG1=   6.00  ARG2=   5.00  X= .96073724E-01  I(X)= 0.53622542224499E-05
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../ 13/
ARG1=   6.00  ARG2=   5.50  X= .39654795E-01  I(X)= 0.10510350814899E-05
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  5/ 12/
ARG1=   6.00  ARG2=   6.00  X= .30561212E-01  I(X)= 0.32967842145126E-06
ARG1=   6.00  ARG2=   6.50  X= .92338593E-01  I(X)= 0.64706134139669E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  3/  5/
ARG1=   6.00  ARG2=   7.50  X= .87443196E-02  I(X)= 0.54050905900271E-09
ARG1=   6.00  ARG2=   7.00  X= .24489806E-01  I(X)= 0.17553590044786E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  3/  4/
ARG1=   6.00  ARG2=   8.00  X= .60971347E-02  I(X)= 0.84995122054623E-10
ARG1=   6.00  ARG2=   7.50  X= .24187674E-01  I(X)= 0.22190408088818E-06
THE FOLLOWING EFFECTS ARE ADMISSIBLE.../  3/
ARG1=   6.00  ARG2=   8.50  X= .60236210E-02  I(X)= 0.10502532177270E-09
ARG1=  13.50  ARG2=   1.00  X= .24903681E-00  I(X)= 0.70722535534782E-08
ARG1=  14.50  ARG2=   1.00  X= .97111423E+00  I(X)= 0.29681323435854E-03
```

Note 6.11

Note 6.12

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 DEPS, 2 DEPENDENT VARIABLES
/ANOVA/BOI/CUMUL /SET 01 Note 6.13

FINAL COMPREHENSIVE ANOVA CUMULATIVE DROPPING Note 6.14

STEP	EFFECT	PRC	I(X)	DIFF MS	DIF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X0302/X03.01		0.67802962	.00021563	00_1	.18101885	.00021563	001	.00119118	012	.99397038
002	01.02X03.02		0.65024463	.00017414	00_1	.16444432	.00019586	00_2	.00119118	012	.99358953
003	01.01X03.02		0.95013573	.00001408	00_1	.11413024	.00013595	003	.00119118	012	.99381129
004	01.02X03.01		0.97235622	.00016785	00_1	.12082549	.00014393	004	.00119118	012	.99389451
005	01.01X03.01		0.99010189	.00002389	00_1	.10067231	.00011992	005	.00119118	012	.99371178
006	01.02X00302		0.05748406	.01982195	00_1	2.8573184	.00340359	006	.00119118	012	.99372371
007	01.01X00302	*	0.01932550	.01195087	00_1	3.8823807	.00462463	007	.00119118	012	.98537068
008	01.02	*	0.00210129	.02997530	00_1	6.5426184	.00779346	008	.00119118	012	.98533455
009	01.01	*	0.00071373	.02381382	00_1	8.0369627	.00957350	009	.00119118	012	.98770288
010	00002X03.02		0.00000105	.13802920	00_2	26.419234	.03147618	011	.00119118	012	.95786767
011	00002X03.01		0.00000065	.04077548	00_2	27.621891	.03290176	013	.00119118	012	.84809854
012	03.02		0.00000018	.14166295	00_1	34.142843	.04067042	014	.00119118	012	.81373274
013	03.01		0.00000022	.00729084	001	32.274698	.03844511	015	.00119118	012	.75403557
014	00002	**	0.00029881	.89102770	00_2	10.888966	.89102770	00_2	.08182849	029	.75096319

EXAMPLE 6...3X3X3 FACTORIAL, 9 CELLS EMPTY, WITH 3 DEPS, 2 DEPENDENT VARIABLES
/ANVA/V1/CUMUL /SET 01

FINAL COMPREHENSIVE ANVA CUMULATIVE DROPPING

Note 6.15

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.01X00002X03.01		0.53770088	.67846154	00_1	.40251509	.67846154	00_1	1.6855556	012	.99757016
002	01.02X03.02		0.72317943	.44415541	00_1	.33301096	.56130847	00_2	1.6855556	012	.99768865
003	01.01X03.02		0.87772538	.01121398	00_1	.22422497	.37794364	00_3	1.6855556	012	.99743530
004	01.02X03.01		0.86286936	.98629243	00_1	.31445463	.93003084	00_4	1.6855556	012	.99743395
005	01.01X03.01		0.92107691	.15280347	00_1	.26969468	.45458537	00_5	1.6855556	012	.99731547
006	01.02X00002	*	0.00221469	68.502583	00_1	6.9982376	11.795918	00_6	1.6855556	012	.99729711
007	01.01X00002	*	0.00039695	44.199812	00_1	9.7445211	16.424932	00_7	1.6855556	012	.98936785
008	01.02		0.00002894	97.207334	00_1	15.739306	26.522732	00_8	1.6855556	012	.98379820
009	01.01		0.00000763	83.643049	00_1	19.500653	32.869434	00_9	1.6855556	012	.97208062
010	00002X03.02	*	0.00000078	445.10355	00_2	29.574839	445.10355	00_2	15.050075	021	.96203253
011	00002X03.01		0.00000070	150.29299	00_2	19.780518	297.69827	00_4	15.050075	021	.65509142
012	03.02		0.00000009	491.55943	00_1	22.356735	336.47050	00_5	15.050075	021	.81898185
013	03.01		0.00000018	22.607144	00_1	18.880965	284.19994	00_6	15.050075	021	.75993052
014	00002	**	0.00028186	3151.6812	00_2	10.979613	3151.6312	00_2	287.04392	029	.75721471

PROBLEM RUNNING TIME(HRS,MIN,SEC)=00:04:11

Note 6.16

3.4.7 Example 7

The purpose of exhibiting Example 7 is to show the capability of NOVACOM in dealing with unacceptable inverses of the matrices (A) of the normal equations. In order to show this, Example 6 was modified such that a singularity is introduced into the matrix A of rank N+1. This was achieved by fitting both confounded constants ab_{12} and bc_{22} while abc_{111} was not fitted, which again makes 17 constants fitted as in Example 6. (In the proper interpretation considering that factors *A* and *C* are quantitative and in the NOVACOM notation, the two confounded constants are 1.1 x 2*2 and 2*2 x 3.2; see also the reproduced input sheet.)

Because of the purpose mentioned, the problem was executed for y_1 only, and as covariates only the 3 OCIVs were used, i.e., no CIVs were generated for Example 7.

The numerical values for y_1 and the 3 OCIVs used in the present case are the same as in Example 6. Again only the FCAs are shown in the reproduced printout.

Notes on printout Example 7.

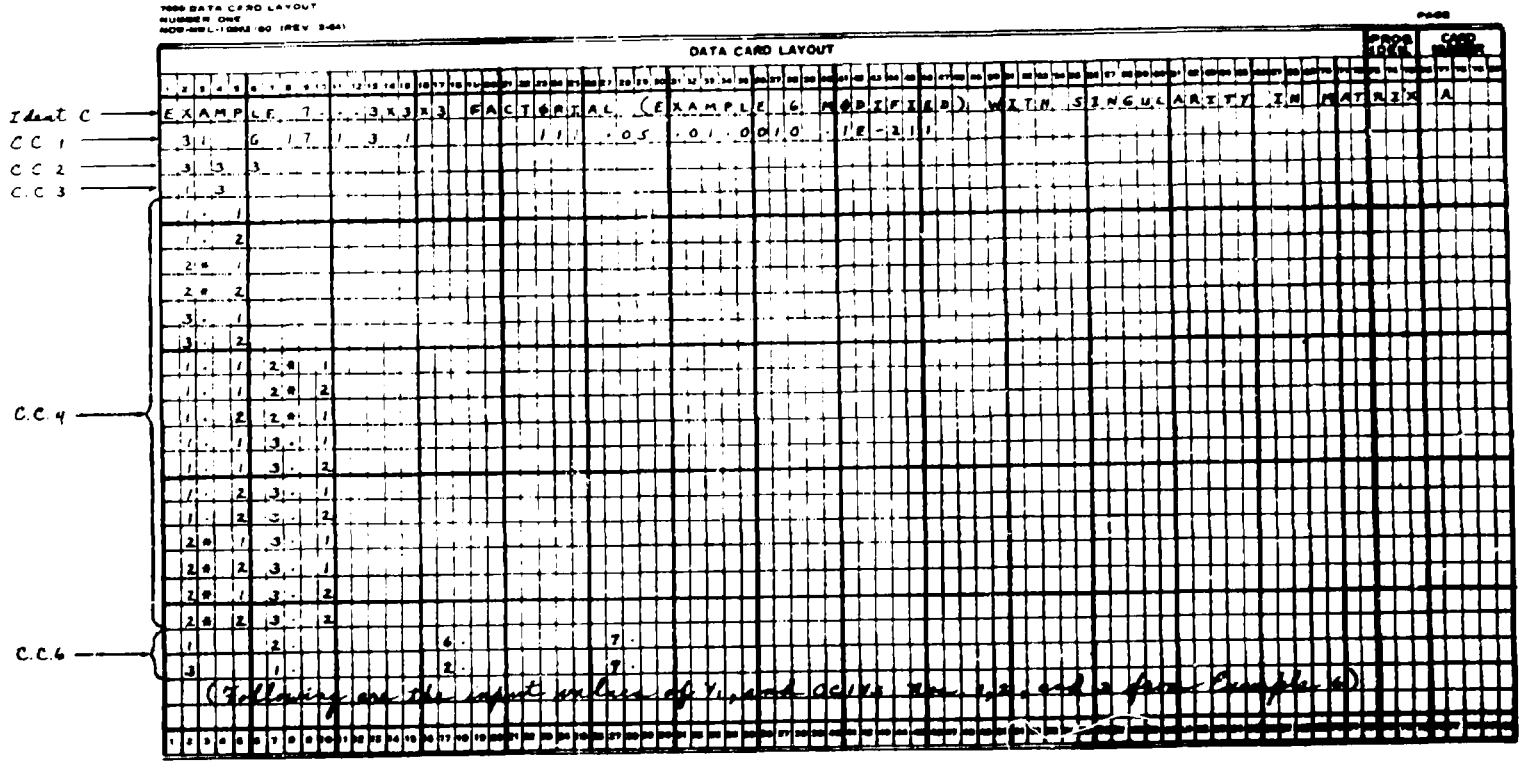
Note 7.1. The inverses, A^{-1}, for all three steps in COMO were rejected because the determinants were found to be negative. The 3 OCIVs were deleted "from the right": in the order of input, OCIV No. 3 was the "rightmost" admissible CIV and, therefore, was deleted from the model at the first step. See also Flowchart No. 4 in Section 2.4.1.

Note 7.2. The FCA for COMO, single dropping, shows the appropriate statement for this case.

Note 7.3. The first 6 steps in FEMO led to rejections of the inverses because only at Step 6 of the ranking was the constant 2*2 x 3.2 deleted from the model whereby the singularity in the matrix of the normal equations was eliminated. Because of the rather arbitrary deletion of effects which are "rightmost" among the effects admissible for ranking at a given step, the program is not very efficient in eliminating the singularity at the earliest possible step. If the effect 2 x 3.2 (which was admissible at the first step!) had been deleted at the first step, the remaining 12 steps would have represented a genuine FEMO ranking. However, the analyst who meets a similar situation will, no doubt, execute the problem a second time after correcting for the cause of the rejections. The results of the first trial will usually be of considerable help to the analyst for the indicated correction. In the present example, the analyst would rightfully suspect that effect 2 x 3.2 caused the previous model rejections and he would take the appropriate corrective action. (The cause for the rejections could be other than in the present example where it was assumed that the analyst made a mistake in fitting the constants. For example, the cause may be the

insufficient accuracy of the inversion process. One may also consider using the corrective capability of NOVACOM for the detection of confounding if it cannot be detected otherwise.) Note that DF(2) = 19 at Step 7 are the 12 degrees of freedom "within cells" pooled with the 7 degrees of freedom due to the 6 deleted effects.

Note 7.4. The FCA of FEMO, single dropping, shows only the "good" steps, i.e., the steps at which the inverses were accepted.



7000 DATA CARD LAYOUT
NUMBER ONE
NOV-NWL-1[illegible]/60 (REV 3-64)

PAGE

DATA CARD LAYOUT

PROG IDEN | CARD NUMBER

Ident C	EXAMPLE 7 . . 3X3X3 FACTORIAL (EXAMPLE 6 MODIFIED) WITH SINGULARITY IN MATRIX A
C C 1	3 1 6 1 7 1 3 1 1 1 1 .05 .01 .0010 .1E-2 1 1
C C 2	3 3 3
C.C 3	1 3
C.C.4	1 . 1
	1 . 2
	2 * 1
	2 * 2
	3 . 1
	3 . 2
	1 . 1 2 * 1
	1 . 1 2 * 2
	1 . 2 2 * 1
	1 . 1 3 . 1
	1 . 1 3 . 2
	1 . 2 3 . 1
	1 . 2 3 . 2
	2 * 1 3 . 1
	2 * 2 3 . 1
	2 * 1 3 . 2
	2 * 2 3 . 2
C.C.6	1 2 . 6 . 7 .
	3 1 . 2 . 7 .
	(Following are the input values of Y1, and CC1 Y2 Nos 1, 2, and 3 from Example 6)

EXAMPLE 7...3X3X3 FACTORIAL (EXAMPLE 6 MODIFIED) WITH SINGULARITY IN MATRIX A
/COMO/V1/CUMUL /DET 1

FINAL COMPREHENSIVE COMO-CUMULATIVE DROPPING Note 7.1

STEP	CONC VAR		PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
1	3	MATRIX COULD NOT BE INVERTED										
2	2(1)	MATRIX COULD NOT BE INVERTED										
3	1(1)	MATRIX COULD NOT BE INVERTED										

EXAMPLE 7...3X3X3 FACTORIAL (EXAMPLE 6 MODIFIED) WITH SINGULARITY IN MATRIX A
/COMO/V1/SINGLE/DET 1

FINAL COMPREHENSIVE COMO-SINGLE DROPPING Note 7.2

STEP	CONC VAR	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET

NO PRINTOUT FOR COMO SINGLE DROPPING SINCE NO VALID SUMS OF SQUARES WERE COMPUTED.

EXAMPLE 7...3X3X3 FACTORIAL (EXAMPLE 6 MODIFIED) WITH SINGULARITY IN MATRIX A
/FEMO/V1/CUMUL /DET 01

FINAL COMPREHENSIVE COMO-CUMULATIVE DROPPING Note 7.3

STEP	EFFECT		PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
001	01.02X03.02	MATRIX COULD NOT BE INVERTED										
002	01.02X03.01	MATRIX COULD NOT BE INVERTED										
003	01.01X03.02	MATRIX COULD NOT BE INVERTED										
004	01.01X03.01	MATRIX COULD NOT BE INVERTED										
005	01.02X00002	MATRIX COULD NOT BE INVERTED										
006	00002X03.02	MATRIX COULD NOT BE INVERTED										
007	00002X03.01			0.93492752	.35869907	002	.06792512	[illegible]	002	5.3113872	019	.98787606
008	01.02			0.46470162	13.447902	001	.88895019	4.7216020	001	5.3113872	019	.98779069
009	03.01			0.33739619	11.638469	001	1.2145309	6.4508437	004	5.3113872	019	.98617923
010	01.01X00002		--	0.00030666	641.81751	002	11.376881	641.81751	002	56.414190	029	.98477708
011	01.02			0.72906120	6.9218786	001	.12269747	6.9218786	001	56.414190	025	.83097324
012	01.01			0.01114283	603.73460	001	5.4122596	305.32824	002	56.414190	029	.82974171
013	00002		--	0.00028186	3151.6912	002	10.979613	3151.6312	002	287.04392	029	.75721471

EXAMP E [illegible] FACTORIAL (EXAMPLE 6 MODIFIED) WITH SINGULARITY IN MATRIX A
/FFM3/Y1/SINGLE/SET [illegible]

FINAL COMPREHENSIVE FFM3-SINGLE DROPPING

Note 7.4

STEP	EFFECT	PRC	I(X)	DIFF MS	DIFF DF	F	MS(1)	DF(1)	MS(2)	DF(2)	COEFF DET
007	[illegible]3.01		0.93492752	.35865207	002	.06752512	.35865207	002	5.3113872	019	.98787686
008	[illegible]3.02		0.11037893	13.447502	001	2.7785828	13.447502	001	4.8394981	021	.98779069
009	[illegible]3.01		0.14999917	11.638569	001	2.2249385	11.638569	001	5.2309819	022	.98617523
010	[illegible]1.01A[illegible]		0.00000000	641.81751	002	116.49175	641.81751	002	5.5095515	023	.98477708
011	[illegible]1.02		0.72906120	6.9218786	001	.12269747	6.9218786	001	56.414190	025	.83057324
012	[illegible]1.01		0.00261845	603.73460	001	11.075937	603.73460	001	54.510639	026	.82974171
013	[illegible]002		0.00000001	3151.6312	002	42.104686	3151.6312	002	74.852268	027	.74721471

PROBLEM RUNNING TIME(HRS,MIN,SEC)=0[illegible]

4. GLOSSARY OF TERMS USED IN THE REPORT

The page numbers in the following alphabetical glossary give the pages where the main definitions are introduced (page number underlined) or where additional pertinent information concerning a term is given (page number not underlined). The glossary is not a complete reference to all pages where a term is discussed or mentioned. Rather, the glossary is intended as a guide to the page where a given term is introduced.

A (= matrix of normal equations) 18

A (= number of levels of factor α) 4

Additional analysis of variance (ANVA) 8,20

"Additional" regression sum of squares (= $SS_{R-R'}$) 10,14,17

Admissibility (of CIV or effect for ranking) 14,15,20

ALPHA (KALPHA) 12

ANVA (= Additional ANalysis of VAriance) 8,20

ASSR(N) (= "total" regression sum of squares adjusted for the mean) 10

ATSS (= total sum of squares adjusted for the mean) 10

Automatic Generation (of CIVs) 21

Automatic Generation (of DIVs) 25

Auxiliary independent variable (u_i) 5

B (= number of levels of factor β) 4

Backward ranking method 9,13

CIV (= Concomitant Independent Variable) 3,21

Coding (of OCIVs) 34

Coding (of quantitative factor variables) 36

COMO (= COncomitant Variables Magnitude [of prediction power for y] Ordering) 9

Complete printout	88
Compound	13
Control Card 4 Set	8,80,91
Cumulative dropping (in COMO)	37
Cumulative dropping (in FEMO)	41
Cumulative ranking	9,13
D (= order of DIV-model)	25
Data input	19
Data matrix	37
Deletion (of CIVs)	23
Deletion (of DIVs)	26
Design Independent Variable (DIV)	3,7
Design matrix	34
DIV (= Design Independent Variable)	3,7
Effect (= factorial effect)	4,26
Factorial effect	4,26
Factor number	25,26
Factor pair	25,27
F_c	10
FCA (= Final Comprehensive Analysis)	20
FEMO (= Factorial Effects Magnitude [of prediction power for y] Ordering)	9
Final Comprehensive Analysis (FCA)	20
Final FCA	20
First good step	19

Full data matrix	66
Full effect	28
Full model	81
Full printout (at significant step of ranking)	18,19
GCIV (= Generated CIV)	3,21
Generated CIV (GCIV)	3,21
Generation (of CIVs)	21
Generation (of DIVs)	25
Good model	19
Good step	19
Hand-generation (of CIVs)	23,83
Hand-generation (of DIVs)	81
I_o (= $A^{-1}A$ = computed identity matrix)	18
ISUBX	12
IV (= Independent Variable)	3
I(X) (= Non-Significance)	12, 20
KALPHA	12
Level number	25
Main Theorem (of multiple regression)	10
Matrix of normal equations (A)	18
"Most probable" significant model	8,91
N (= total number of independent variables)	3
n (= total number of observed y-values)	10

"Non-orthogonal" analysis of variance 1
Non-Significance (I(X)) 11
NOVACOM 1
OCIV (= Original CIV) 3,20
OCIV number 21
Order (of DIV) 23
Original CIV (OCIV) 3,20
"Orthogonal" analysis of variance 1
P (= order of CIV-model) 21
Partially Fitted Full Effect (PFFE) 17,29
PFFE (= Partially Fitted Full Effect) 17,29
Power (of CIV) 21,23
Power (of quantitative factor variable) 25
Power-sum 82
$^+$-procedure 18,45
Qualitative factor 4
Quantitative factor 6
Quantitative factor variable ($X_\alpha, X_\beta, \ldots$) 6
$R_{\alpha\beta\gamma\ldots}$ (= number of observations in cell $\alpha\beta\gamma\ldots$) 4
Rejected model 19
Relaxed admissibility (of effects for ranking) 17,28
Restricted admissibility (of CIVs or effects for ranking) 14,15,16,17,23,27
Restriction-dependence (of $SS_{N-N'}$-values) 14,17
Significant model 9,14,15,18,19

Single dropping	13
Single dropping (in COMO)	40
Single dropping (in FEMO)	49
$SS_{N-N'}$ (= "additional" regression sum of squares)	10,14,17
Sub-CIV	16,23
Sub-DIV	16,27
Sub-effect	14,27
Summation matrix	19,37
T (= total number of covariates)	3,21
TOLI2	18
"Total" regression sum of squares (= ASSR(N))	10
TP (= total number of OCIVs)	21
u_ν (= ν^{th} auxiliary independent variable)	5,7
Unrestricted admissibility (of CIVs or effects for ranking)	17
W (= number of dependent variables in a problem)	66
X_α (= quantitative factor variable for factor α)	6
X_β (= quantitative factor variable for factor β)	6
Zero error perfect fit	68

5. REFERENCES

1. Abt, K. [1960]. Analyse de Covariance et Analyse par Différences. Metrika 3, 26-45, 95-116, 177-211.

2. Abt, K., Gemmill, G., Herring, T., and Shade, R. [1966]. DA-MRCA: A FORTRAN IV Program for Multiple Linear Regression. U. S. Naval Weapons Laboratory Technical Report No. 2035.

3. Abt, K. [1967]. On the Identification of the Significant Independent Variables in Linear Models. Metrika 12, 1-15.

4. Anderson, R. L. and Bancroft, T. A. [1952]. Statistical Theory in Research. McGraw-Hill Book Co., Inc., New York.

5. Brownlee, K. A. [1960]. Statistical Theory and Methodology in Science and Engineering. John Wiley and Sons, Inc., New York.

6. Davies, O. L. (Ed.) [1956]. The Design and Analysis of Industrial Experiments. Hafner Publ. Co., New York.

7. DiDonato, A. R. and Jarnagin, M. P. [1960]. A Method for Computing the Incomplete Beta Function Ratio. U. S. Naval Weapons Laboratory Technical Report No. 1949, revised.

8. Duncan, A. J. [1959]. Quality Control and Industrial Statistics. Richard D. Irwin, Inc., Homewood, Illinois.

9. Gosslee, D. G. and Lucas, H. L. [1965]. Analysis of Variance of Disproportionate Data when Interaction is Present. Biometrics 21, 115-133.

10. Graybill, F. A. [1961]. An Introduction to Linear Statistical Models, Volume I. McGraw-Hill Book Co., Inc., New York.

11. Greenwood, J. A. and Hartley, H. O. [1962]. Guide to Tables in Mathematical Statistics. Princeton University Press, Princeton, New Jersey.

12. Hader, R. J. and Grandage, A. H. E. [1958]. Simple and Multiple Regression Analyses. In: Experimental Designs in Industry. (Chew, V., Ed.) John Wiley and Sons, Inc., New York.

13. Herring, T. [1968]. A Programming Guide to NOVACOM, A [illegible] Computer Program for "Non-orthogonal" Analysis of Variance and Covariance. U. S. Naval Weapons Laboratory Technical Report No. TR-2157.

14. Hicks, C. R. [1964]. Fundamental Concepts in the Design of Experiments. Holt, Rinehart and Winston, New York.

15. Robson, D. S. [1959]. A Simple Method for Constructing Orthogonal Polynomials when the Independent Variable is Unequally Spaced. Biometrics 15, 187-191.

16. Scheffé, H. [1959]. The Analysis of Variance. John Wiley and Sons, Inc., New York.

17. Stevens, W. L. [1948]. Statistical Analysis of a Non-Orthogonal Tri-Factorial Experiment. Biometrika 35, 346-367.

APPENDIX A

BLANK PAGE

Appendix A:

METHOD OF FITTING CONSTANTS FOR NON-ORTHOGONAL LAYOUTS WITH INTERACTIONS AND EMPTY CELLS

The method proposed in this Appendix is developed for the case of only qualitative factors in a given data layout. However, an extension to cases with quantitative factors is easily possible. The method of fitting constants is treated strictly from the viewpoint of hypothesis testing. Therefore, emphasis is put on the proofs that the null hypotheses are testable when the backward ranking technique of the factorial effects is applied.

In order to introduce some of the concepts of the proposed method, the two-way crossed classification example from Section 2.1.1 of the present report is used (Example A below) together with two modifications (Examples B and C). Then, a three-way crossed classification example (Example D) where all cells are occupied is treated. Finally, all the essential features of the method are exemplified, in a combined manner, with a three-way crossed classification where some cells are empty (Example E).

Example A

In Figure 1 the layout of Example A is given together with the two marginal one-way classifications for factors α and β. The following are some of the concepts and symbols which are used for the various features of the fitting process: A cell is identified by the sequence of the factor level symbols (in alphabetical order of the factors) which uniquely define the cell. For example, the cell identification for the cell defined by the first level of α and the second level of β is given by $\alpha_1\beta_2$. A distinction is made between "basic cells" and "marginal cells": basic cells are those of the basic (original) classification, whereas marginal cells are those of the marginal classifications which result from summing over all levels of at least one factor. For example, cell $\alpha_1\beta_2$ in Figure 1 is a basic cell, and cell α_1 is a marginal cell (resulting from summing over the 3 levels of factor β). A "row" of cells is defined as the group of cells (basic or marginal) which is formed by keeping constant the levels of all but one factor in the layout of the basic or marginal cells. For example, by keeping $\beta = 1$ constant in Figure 1, the three cells $\alpha_1\beta_1$, $\alpha_2\beta_1$, and $\alpha_3\beta_1$ form a row of basic cells. An "X" in a cell (basic or marginal) means that the cell is occupied, i.e., that there are observations in this cell. A circle around the "X" means that a constant (parameter) has been fitted based on the observation(s) in this cell; and a checkmark (of one of three types to be defined) through an "X" means that a constant either has not been fitted or can not be fitted based on the observations in the checkmarked cell.

In Figure 1 the two marginal one-way classifications ("rows of marginal cells" by structure) will be used to demonstrate the fitting of main effect constants, and the two-way classification (of the basic cells) will be used to demonstrate the fitting of the interaction constants.

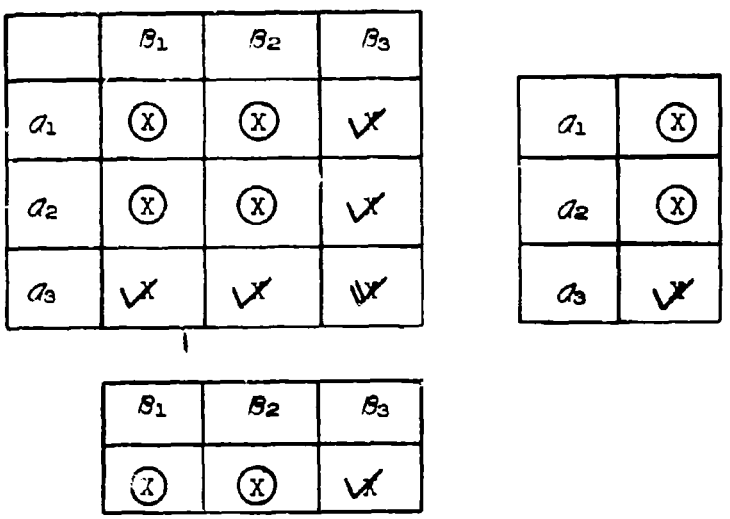


Figure 1: Layout of Example A.

Legend:

✓ : Type I-Checkmark
(See Rule I)

✓ : Type II-Checkmark
(See Rule IIa)

Since a linear restriction has to be imposed on each set of main effect constants (i.e., for α and β), a constant based on the observations in one of the (A=3 and B=3) marginal cells is linearly dependent upon the constants based on the observations in the other (two) cells. Generally, the two constants do not have to be based on the first two cells (as done for both factors α and β in Figure 1). However, if the linear restrictions of the Graybill type, $a_A = b_B = 0$, are chosen as suggested in Section 2.1.1 of the present report, the last cell can not be used as a basis for fitting a constant since this constant is eliminated a priori from the model.

Under the corresponding restrictions for the interaction constants, ($ab_{\alpha\beta} = ab_{A\beta} = 0$ for $\alpha = 1,\ldots,A$; $\beta = 1,\ldots,B$) the same argument applies to each row and each column of the two-way classification: only the interaction constants ab_{11}, ab_{12}, ab_{21}, and ab_{22} can be fitted. This completes the full set of AB-1=8 constants contained in the model as given in equation (2-5) of Section 2.1.1.

The argumentation just used is the basis for the fitting of main effect and interaction constants by visual inspection, according to which the fitting will be performed from here on: circles are used (Figure 1) in the four cells $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{B}_1$, and $\mathcal{B}_2$ to indicate that main effect constants have been based on the observations in these cells. As a consequence of this choice, the last cells in both rows of marginal cells have to be checkmarked. "Checkmarking" is used here as a synonym for equating to zero the constant which would have been based on the cell if it had been possible. The first rule for checkmarking a cell is thus stated:

> Rule I. If, in a row of basic or marginal cells, all but one occupied cell have been circled (where the choice of cells to be circled is up to the analyst), the one occupied cell left will receive a "Type I - checkmark": see legend in Figure 1.

Note that, as indicated before, it would have been possible to choose, for example, the marginal cells $\mathcal{A}_1$ and $\mathcal{A}_3$ for circling, which would have left cell $\mathcal{A}_2$ to receive a Type I-checkmark. (This would have implied a set of linear restrictions different from the Graybill type.)

In fitting the four interaction constants in the two-way layout of basic cells as shown in Figure 1, the last cells in the 4 rows defined by α=1, α=2, β=1, and β=2 also receive Type I-checkmarks according to Rule I. Then, the only occupied cell not yet considered in the fitting process by visual inspection is $\mathcal{A}_3\mathcal{B}_3$. The reasoning for not being able, in the visual process, to base an interaction constant on this cell, given the four interaction constants have been fitted as indicated in Figure 1, is as follows. Each interaction effect (of any order), to be represented by a fitted constant, must be interpretable as a contrast of contrasts and, therefore, requires two occupied cells in the rows of (basic or marginal) cells with which the effect is to be associated. Further, if two occupied cells in a row are available, a choice must obviously exist to actually base the constant on the one or the other occupied cell. This feature of having the choice to base the constant on either one of the cells in the row will be called "reversibility."

Inspecting, in Figure 1, the row of basic cells defined by β=3, one can see that once the cells $\mathcal{A}_1\mathcal{B}_3$ and $\mathcal{A}_2\mathcal{B}_3$ are checkmarked (by Rule I), there are no cells left (in the row $\mathcal{B}$=3) for which reversibility exists. Therefore, no interaction constant can be based on cell $\mathcal{A}_3\mathcal{B}_3$, and it is checkmarked according to "Rule IIa":

Rule IIa. If, in a row of basic or marginal cells, all but one cell have been checkmarked according to Rule I such that no reversibility (as defined above) exists for the one cell left, the cell will receive a "Type II-checkmark": see legend in Figure 1.

(Note: A "Rule IIb" and a "Rule IIc" according to which the Type II-checkmark will again be applied, are given in the discussions of Examples B and D, respectively, after the definition of a "Rule III.")

Notice that, with respect to cell $\mathcal{A}_3\mathcal{B}_3$, the necessity of applying the Type II-checkmark is also evidenced by the two Type I-checkmarks in the row defined by $\alpha=3$.

This completes the fitting process by visual inspection, following the established Rules I and IIa, for Example A.

Performing, in Example A, the analysis of variance corresponding to the backward ranking process under restricted admissibility (see Section 2.1.2), the only admissible null hypothesis at the first step is the null hypothesis concerning the interaction effect $\mathcal{AB}$, i.e., $H_0\{ab_{11}=ab_{12}=ab_{21}=ab_{22}=0\}$. This joint hypothesis is testable since for each of the four ab-constants there is a linear function of the observations having the particular ab-constant (parameter) as expectation under the given model. For example, $E[y_{12\rho}-y_{13\rho}-y_{32\rho}+y_{33\rho}] = ab_{12}$. At the second step, provided the null hypothesis about $\mathcal{AB}$ was not rejected and the ab-constants were deleted from the model, the null hypotheses about the main effects of both factors $\mathcal{A}$ and $\mathcal{B}$ are admissible. Both hypotheses $H_0\{a_1=a_2=0\}$ and $H_0\{b_1=b_2=0\}$ are testable since there are linear functions of the observations which have a_1, a_2, b_1, or b_2, whichever is applicable, as expectation under the given model.

Considering only degrees of freedom, the AB-1=8 degrees of freedom "between cells" are assigned to the three factorial effects as follows, as would be expected for a layout in which all cells are occupied:

$\mathcal{A}$	2
$\mathcal{B}$	2
$\mathcal{AB}$	4
Total = "between cells"	8

Example B

Example B is a modification of the previous Example A: the (basic) cells $\alpha_1\beta_2$, $\alpha_2\beta_3$, and $\alpha_3\beta_3$ are now empty, whereas the other six cells are occupied as before; see Figure 2. By employing the fitting process by visual inspection, the main effect constants are fitted as before which consumes 4 of the now 5 available degrees of freedom "between cells." Obviously, this time only 1 interaction constant can be fitted.

Inspection of the two-way layout in Figure 2 shows that cell $\alpha_1\beta_3$ is the only occupied cell in the row of basic cells defined by $\beta=3$. As stated before, any fitted constant requires the availability of two occupied cells in the row of cells with which the effect (= contrast), represented by the fitted constant, is to be associated. Therefore, certainly no constant can be based on a single occupied cell in a row such as $\alpha_1\beta_3$ in Example B. This leads to the simple "Rule III":

Rule III. If, in a row of basic or marginal cells, there is only one cell occupied, this cell will receive a "Type III-checkmark": see legend in Figure 2.

	β_1	β_2	β_3
α_1	✓		✓
α_2	⊗	✓	
α_3	✓	✓	

α_1	⊗
α_2	⊗
α_3	✓

β_1	β_2	β_3
⊗	⊗	✓

Figure 2: Layout of Example B.

Legend:

✓ : Type II-Checkmark
(See Rules IIa and IIb)

✓ : Type III-Checkmark
(See Rule III)

Once cell $\alpha_1\beta_3$ in Figure 2 is checkmarked, cell $\alpha_1\beta_1$ remains the only occupied cell not checkmarked in row α=1. Therefore, no reversibility exists for the two occupied cells in that row, and it is not possible to base an interaction constant on cell $\alpha_1\beta_1$ either. However, unlike in Example A, this time the checkmarking of a cell, which is the only cell not yet checkmarked in a row of cells, is not a consequence of fitting constants by choice (i.e., of circling cells in other rows), but a consequence of checkmarking a cell when no alternative exists. This leads to the definition of "Rule IIb":

> Rule IIb. If, in a row of basic or marginal cells, all but one cell have been checkmarked according to Rule III, the one cell left will receive a "Type II-checkmark." (See legend in Figure 2.)

The fitting process by visual inspection in Example B is completed by choosing to base the one interaction constant on cell $\alpha_2\beta_1$, for example. This is done in Figure 2, and cell $\alpha_2\beta_1$ is circled accordingly. As a consequence, cells $\alpha_2\beta_2$ and $\alpha_3\beta_1$ are checkmarked following Rule I and cell $\alpha_3\beta_2$ is checkmarked following Rule IIa.

At the first step of the ranking process, the only admissible null hypothesis is again that on the interaction, $H_0\{ab_{21}=0\}$. This hypothesis is testable since $E[y_{21\rho}-y_{22\rho}-y_{31\rho}+y_{32\rho}] = ab_{21}$ under the given model.

Example C

The layout of Example C, as given in Figure 3, results from Example A by deleting the observations in the four cells $\alpha_1\beta_2$, $\alpha_2\beta_1$, $\alpha_2\beta_3$, and $\alpha_3\beta_2$, as indicated.

	β_1	β_2	β_3
α_1	(X)		X
α_2		X	
α_3	X		X

α_1	(X)
α_2	(X) I
α_3	X

β_1	β_2	β_3
(X)	(X) I	X

Figure 3: Layout of Example C.

Now five basic cells are occupied and it appears, from applying Rules III, I, and IIa, (see the checkmarks in Figure 3), that five constants can be fitted as is indicated by the circles. (The two "I"s in the marginal cells $\mathcal{A}_2$ and $\mathcal{B}_2$ will be explained later.) However, actually only 4 degrees of freedom "between cells" are available to be assigned to factorial effects. The reason for the discrepancy is simple: The deletion of the observation(s) in cell $\mathcal{A}_2\mathcal{B}_2$ would lead not only to the loss of one degree of freedom "between cells", but also to the loss of one degree of freedom for each of the main effects of factors $\mathcal{A}$ and $\mathcal{B}$. In other words, the observation(s) occupying the basic cell $\mathcal{A}_2\mathcal{B}_2$ cause both marginal cells $\mathcal{A}_2$ and $\mathcal{B}_2$ to be occupied; that is , deleting all observations from cell $\mathcal{A}_2$ also deletes all observations from cell $\mathcal{B}_2$. The type of relation among non-empty cells (basic or marginal) thus exemplified will be expressed in an algebraic identity containing the symbols (identifications) of the cells involved in the relation, that is, in the present example

$$\mathcal{A}_2 \equiv \mathcal{B}_2.$$

Each such identity represents one confounded degree of freedom. That is, one degree of freedom of all those factorial effects, whose constants are based on the cells represented by their cell symbols in the identity, is confounded. In Example C, therefore, one degree of freedom of each of the main effects is confounded since the constants a_2 and b_2 are based on the (marginal) cells $\mathcal{A}_2$ and $\mathcal{B}_2$, respectively. For this reason, cells $\mathcal{A}_2$ and $\mathcal{B}_2$ are marked with an "I" (for Identity) in Figure 3. Clearly in this case, either a_2 or b_2 can be fitted, but not both simultaneously. Therefore, at the first step of ranking in this example, $H_0\{ab_{11}=0\}$ will be tested with the model containing either the constants a_1, a_2, b_1, or a_1, b_1, b_2. At this first step it makes no difference whether a_2 or b_2 is fitted in addition to a_1 and b_1.

Once the interaction constant ab_{11} is deleted from the model (assuming that $H_0\{ab_{11}=0\}$ was not rejected) the fitting of a_2 or b_2 depends upon which null hypothesis is to be tested. For example, in order to test $H_0\{b_1=0\}$, i.e., to test the hypothesis that there is no main effect due to factor $\mathcal{B}$ in addition to the main effect of factor $\mathcal{A}$, the reduced model must contain the constants a_1 and a_2. The corresponding argument holds for the testing of $H_0\{a_1=0\}$, in which case b_1 and b_2 must be contained in the reduced model. Both null hypotheses about a_1 and b_1 are testable, since, for example, $E[y_{11\rho}-y_{13\rho}] = b_1$ and $E[y_{11\rho}-y_{21\rho}] = a_1$ under the models containing a_1, a_2, b_1, and a_1, b_1, b_2, respectively. Naturally, not rejecting the hypothesis $H_0\{b_1=0\}$, for example, does not imply that the overall main effect of $\mathcal{B}$ is not significant, but it does imply, given the pattern of empty cells, that the differences among the 5 cell means can sufficiently be explained by the main effect of factor $\mathcal{A}$ alone. In this case, once the constant b_1 has

been deleted from the model, the hypothesis $H_0\{a_1=a_2=0\}$ is testable. A similar argument holds for the case of not rejecting the hypothesis $H_0\{a_1=0\}$. Rejecting $H_0\{a_1=0\}$ or $H_0\{b_1=0\}$ means that the corresponding main effect is significant, at least based on the one unconfounded degree of freedom.

The concept of identities will be further discussed in Example E below.

Example D

Example D is the basis for Example E which will be used to demonstrate all features of the fitting process in a combined manner. Example D results from Example A by introduction of a third factor, C, with C=2 levels. All AxBxC = 3x3x2 cells are assumed occupied, and the layout is given in Figure 4.

		B_1	B_2	B_3
a_1	c_1	(X)	(X)	X
	c_2	X	X	X
a_2	c_1	(X)	(X)	X
	c_2	X	X	X
a_3	c_1	X	X	X
	c_2	X	X	X

Figure 4: Layout of Example D.

The three marginal one-way classifications and the three marginal two-way classifications are not shown since the fitting of constants and testing of null hypotheses for main effects and first order interactions correspond to those shown in Example A.

The fitting of second-order interaction constants (abc-terms) by visual inspection follows the rules established before. For example, in the row of basic cells defined by $\beta=2$ and $\gamma=1$ only two abc-constants can be fitted. Circling the cells $\mathcal{A}_1\mathcal{B}_2\mathcal{C}_1$ and $\mathcal{A}_2\mathcal{B}_2\mathcal{C}_1$ (i.e., fitting the constants abc_{121} and abc_{221}) leads to the checkmarks in cells $\mathcal{A}_1\mathcal{B}_2\mathcal{C}_2$, $\mathcal{A}_2\mathcal{B}_2\mathcal{C}_2$, and $\mathcal{A}_3\mathcal{B}_2\mathcal{C}_1$ according to Rule I. The other two constants fitted are abc_{111} and abc_{211}, and the checkmarks in the remaining cells except $\mathcal{A}_3\mathcal{B}_3\mathcal{C}_2$ are applied in an obvious manner following Rules I and IIa. The checkmark (of Type II) in cell $\mathcal{A}_3\mathcal{B}_3\mathcal{C}_2$ is applied following a similar reasoning as that used for Rule IIa: Cell $\mathcal{A}_3\mathcal{B}_3\mathcal{C}_2$ is the only occupied cell left unmarked in all three rows of basic cells to which it belongs. In these three rows (defined by $\alpha=3$, $\beta=3$; $\alpha=3$, $\gamma=2$; and $\beta=3$, $\gamma=2$) all other cells have been checkmarked according to Rule IIa as a consequence of previous checkmarking according to Rule I which was done as a consequence of fitting the four abc-constants as indicated. Therefore, no reversibility exists for the remaining cells once the 4 cells as indicated are circled, and cell $\mathcal{A}_3\mathcal{B}_3\mathcal{C}_2$ accordingly is also checkmarked. This argumentation can also be generalized to higher-way layouts and thus leads to the last rule to be defined for the fitting process by visual inspection:

> Rule IIc. If, in a row of basic or marginal cells, all but one cell have been checkmarked according to Rule IIa, or, as a consequence of Rule IIa, according to the present Rule IIc, such that no reversibility (as defined before) exists for the occupied cells of the row concerned, the one cell will receive a "Type II-checkmark."

Note. Rules IIa, IIb, and IIc could be combined into one "Rule II" which would state the following: Any cell will receive a Type II-checkmark which is left as the only unmarked occupied cell in a row of basic or marginal cells where all other cells have been checkmarked according to Rule I or III or, as a consequence of Rule I or III, according to "Rule II."

After finishing the fitting process for Example D, the choice of the four abc-constants fitted can be seen to correspond to the choice of the linear restrictions of the Graybill type. The restrictions read, for the three-factor interaction constants: $abc_{\alpha\beta c} = 0$ for all (α,β); $abc_{\alpha\beta\gamma} = 0$ for all (α,γ); and $abc_{\alpha\beta\gamma} = 0$ for all (β,γ).

At the first step of the ranking process, the only null hypothesis admissible is that concerning the three-factor interaction $\mathcal{ABC}$, and obviously, this hypothesis is testable.

Example E

This example results from Example D by deletion of the observations in 5 cells as indicated in Figure 5a. The example contains all the essential

features of the method as they were successively introduced in Examples A through D. Example E, therefore, will be used to demonstrate all the essential aspects of the proposed method.

In addition to the layout of the three-way classification given in Figure 5a (in which factor C is treated as a "subclassification"), Figures 5b and 5c show the same classification arranged such that factors B and A, respectively, are treated as the "subclassification." These three possible arrangements are convenient for the demonstration as will be seen. Figures 5d, 5e, and 5f show the three marginal two-way classifications for fitting AB-, AC-, BC-interaction constants, respectively.

		B_1	B_2	B_3
a_1	c_1	(X)		
	c_2			
a_2	c_1			
	c_2			
a_3	c_1			
	c_2			

Figure 5a: Layout of Example E. (C as "subclassification")

		$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$
$\mathcal{C}_1$	$\mathcal{B}_1$	X	X	X
	$\mathcal{B}_2$		X	
	$\mathcal{B}_3$	X		X
$\mathcal{C}_2$	$\mathcal{B}_1$	X		X
	$\mathcal{B}_2$		X	X
	$\mathcal{B}_3$	X	X	X

Figure 5b: Example E. ($\mathcal{B}$ as "subclassification")

		$\mathcal{C}_1$	$\mathcal{C}_2$
$\mathcal{B}_1$	$\mathcal{A}_1$	X	X
	$\mathcal{A}_2$	X	
	$\mathcal{A}_3$	X	X
$\mathcal{B}_2$	$\mathcal{A}_1$		
	$\mathcal{A}_2$	X	X
	$\mathcal{A}_3$		X
$\mathcal{B}_3$	$\mathcal{A}_1$	X	X
	$\mathcal{A}_2$		X
	$\mathcal{A}_3$	X	X

Figure 5c: Example E. ($\mathcal{A}$ as "subclassification")

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$
$\mathcal{A}_1$	(X)		✓
$\mathcal{A}_2$	(X) I	(X)	✓
$\mathcal{A}_3$	✓	✓	✓

Figure 5d: Example E, Interaction $\mathcal{AB}$

	$\mathcal{C}_1$	$\mathcal{C}_2$
$\mathcal{A}_1$	(X)	✓
$\mathcal{A}_2$	(X) I	✓
$\mathcal{A}_3$	✓	✓

Figure 5e: Example E, Interaction $\mathcal{AC}$

	$\mathcal{C}_1$	$\mathcal{C}_2$
$\mathcal{B}_1$	(X)	✓
$\mathcal{B}_2$	(X) I	✓
$\mathcal{B}_3$	✓	✓

Figure 5f: Example E, Interaction $\mathcal{BC}$

A convenient first approach is to fit, by the process of visual inspection, all constants which appear as if they can be fitted, according to the rules previously established, in all marginal classifications and in the basic (three-way) classification. If, in this way, k more constants result than there are degrees of freedom "between cells", then exactly k identities must exist for the given data. Naturally, if the number of constants thus fitted is equal to the number of degrees of freedom "between cells", identities do not exist for the given data. In the present example, the fitting of main effect constants (the three marginal one-way classifications not shown for this example) and first-order interaction constants (see Figures 5d-5f) yields 2, 2, and 1 constants for $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, respectively; and 3, 2, and 2 constants for $\mathcal{AB}$, $\mathcal{AC}$, and $\mathcal{BC}$, respectively. In order to evaluate the possibilities of fitting abc-constants, consider Figure 5a. Four cells, as indicated, are checkmarked according to Rule III. For example, cell $\mathcal{A}_2\mathcal{B}_2\mathcal{C}_1$ is the only occupied cell in the row of basic cells defined by $\beta=2$ and $\gamma=1$. Cell $\mathcal{A}_2\mathcal{B}_2\mathcal{C}_2$ is checkmarked following Rule IIb (as a consequence of the Type III-checkmark in cell $\mathcal{A}_2\mathcal{B}_2\mathcal{C}_1$). Without fitting an abc-constant first, the remaining 8 occupied cells can not be checkmarked. If cell $\mathcal{A}_1\mathcal{B}_1\mathcal{C}_1$ is chosen as the basis for a constant fitted (i.e., for abc_{111}), the remaining 7 cells can be checkmarked following Rules I, IIa, and IIc. Summarizing the results from the fitting process, the following degrees of freedom are preliminarily assigned to the 7 factorial effects:

$\mathcal{A}$	2
$\mathcal{B}$	2
$\mathcal{C}$	1
$\mathcal{AB}$	3
$\mathcal{AC}$	2
$\mathcal{BC}$	2
$\mathcal{ABC}$	1
Total	13

However, only 13 basic cells are occupied (Figure 5a), consequently, there are only 12 degrees of freedom "between cells." Accordingly, 13-12=1 identity must be present in the given layout.

The search for the identities, if these are not obvious as in Example C, can be done either systematically or by trial and error. For example, the analyst can systematically delete the observations of each basic and marginal cell, cell by cell, and examine the numbers of constants which can be fitted in each situation. In general, this examination should give sufficient hints as to the "location" of the identities. On the other hand, some experience with the peculiarities of identities will enable the analyst to find the identities of a given layout much faster by trial and error. In the present example, for instance, it is not difficult to find that deleting the observations from the marginal cells $\mathcal{A}_1\mathcal{B}_1$ and $\mathcal{B}_2\mathcal{C}_1$ causes

the marginal cell $\alpha_2\gamma_1$ to be empty too. Algebraically, this relationship is expressed by the identity

$$\alpha_2\beta_1 + \beta_2\gamma_1 \equiv \alpha_2\gamma_1 .$$

This identity implies the following (see Figures 5d-5f in which the 3 affected cells are marked with an "I"): One degree of freedom is confounded in each of the three two-factor interactions, i.e., in $\alpha\beta$, $\alpha\gamma$, and $\beta\gamma$. The three constants affected are those based on the observations in the three marginal cells whose symbols are contained in the identity, i.e., ab_{21}, ac_{21}, and bc_{21}. Any pair of these three constants can be fitted; the simultaneous fitting of all three constants is not possible since it would lead to a singular matrix of the normal equations.

The above identity happens to contain only cell symbols for which no factor is at its last level, i.e., at α=A=3, β=B=3, or γ=C=2. This absence of last levels is very desirable since cell symbols at the last level of any factor can not be associated with constants to be fitted because these constants are deleted a priori from the model if one uses the suggested linear restrictions of the Graybill type. Whenever applicable and possible, therefore, the levels of the factors should be interchanged such that the identities contain only cell symbols in which none of the factors is at its last level. The interchanging is feasible when the method being discussed is applied to cases with only qualitative factors or (quantitative) factors which are treated as qualitative factors. Should it be impossible to free the identities from cell symbols at last factor levels, it will still be possible to find, for each identity, a set of constants which can not be fitted simultaneously. (Note that, for the proper testing of null hypotheses, a set of constants which can not be fitted simultaneously must be found.) The search for this set of constants again may have to be done by trial and error, i.e., by observing whether or not the matrix of the normal equations is non-singular while trying various possibilities of fitting.

There may be more than one identity for a given set of data. However, all identities must be linearly independent from each other in order to account for one confounded degree of freedom each. (The latter implies that, in a system of identities, the cell symbols can be added to and subtracted from each other, which is stated without proof.) For instance, two or more identities are linearly dependent when they can be added to yield a "trivial" identity. A trivial identity is one which does not account for a confounded degree of freedom. In the present Example E one such trivial identity is (see Figure 5a):

$$\alpha_2\beta_2 + \alpha_3\beta_2 \equiv \beta_2\gamma_1 + \beta_2\gamma_2$$

As can be seen, with cell $\mathcal{A}_1\mathcal{B}_2$ being empty, both sides of the identity are equal to $\mathcal{B}_2$.

After finding the identities (if present) and finding, for each identity, one set of confounded constants which can not be fitted simultaneously, the appropriate null hypotheses can be tested.

In the present example, at the first step of the ranking process, only $H_0\{abc_{111}=0\}$ is admissible for testing (assuming there is a possibility for testing, i.e., there exists a valid estimate of the experimental error). This hypothesis is testable since there is a linear function of the observations which has abc_{111} as expectation, under the model containing the constants abc_{111}, ab_{11}, ab_{22}, ac_{11}, bc_{11}, a_1, a_2, b_1, b_2, c_1, plus any two of the three confounded interaction constants, ab_{21}, ac_{21}, bc_{21}:

$$E[(y_{111\rho}-y_{131\rho}-y_{112\rho}+y_{132\rho}) - (y_{311\rho}-y_{331\rho}-y_{312\rho}+y_{332\rho})] = abc_{111}.$$

Assuming $abc_{111}=0$ to be true (and, consequently, assuming abc_{111} to have been deleted from the model), at the second step of the ranking process for this example, the null hypotheses about the interaction effects $\mathcal{AB}$, $\mathcal{AC}$, $\mathcal{BC}$ are admissible for testing. In order to see how the confounding will affect the possibilities of testing, a list of the expected values of the functions indicated below is advantageous. For the present investigation only, the model is assumed to contain all seven interaction constants (besides the five main effect constants), that is ab_{11}, ab_{21}, ab_{22}, ac_{11}, ac_{21}, bc_{11}, and bc_{21}.

The construction of the functions D_j (of individual observations, $y_{\alpha\beta\gamma\rho}$), having the desired expected values, was facilitated by inspection of Figures 5a, 5b, and 5c:

$$E[D_1] = E[y_{111\rho}-y_{311\rho}-y_{131\rho}+y_{331\rho}] = ab_{11}$$

$$E[D_2] = E[y_{222\rho}-y_{322\rho}-y_{232\rho}+y_{332\rho}] = ab_{22}$$

$$E[D_3] = E[y_{111\rho}-y_{112\rho}-y_{311\rho}+y_{312\rho}] = ac_{11}$$

$$E[D_4] = E[y_{111\rho}-y_{131\rho}-y_{112\rho}+y_{132\rho}] = bc_{11}$$

$$E[D_5] = E[y_{221\rho}-y_{222\rho}-y_{331\rho}+y_{332\rho}] = ac_{21}+bc_{21}$$

$$E[D_6] = E[y_{111\rho}-y_{211\rho}-y_{132\rho}+y_{232\rho}] = ab_{11}+ac_{11}-ab_{21}-ac_{21}$$

$$E[D_7] = E[y_{221\rho}-y_{211\rho}-y_{322\rho}+y_{312\rho}] = ab_{22}-bc_{11}-ab_{21}+bc_{21}$$

The last two functions will be replaced by linear combinations with other functions such that the expected values of the new functions contain only confounded effects:

$$E[D_6'] = E[D_1+D_3-D_6] = ab_{21}+ac_{21}$$

$$E[D_7'] = E[D_2-D_4-D_7] = ab_{21}-bc_{21}$$

When equating to zero one of the three confounded constants, i.e., ab_{21}, ac_{21}, or bc_{21}, one can see that the null hypotheses on all three interaction effects, $\mathcal{AB}$, $\mathcal{AC}$, and $\mathcal{BC}$, are testable. For example, if ab_{21} is set equal to zero (i.e., if ab_{21} is deleted from the model), the null hypotheses $H_0[ab_{11}-ab_{22}=0]$, $H_0[ac_{11}-ac_{21}=0]$, and $H_0[bc_{11}-bc_{21}=0]$ are testable since the functions D_1, D_2, D_3, D_6', D_4, and $(-D_7')$ have the respective constants as expectations. However, if according to the test result, $H_0[ab_{11}-ab_{22}=0]$ does not have to be rejected and is assumed to be valid, one does not have evidence that the interaction $\mathcal{AB}$ is not significant. The only valid conclusion is, given the pattern of empty cells, that all two-factor interaction effects (if present) can sufficiently be explained by the interactions $\mathcal{AC}$ and $\mathcal{BC}$. Corresponding arguments apply when ac_{21} or bc_{21} are deleted from the model and $H_0[ac_{11}=0]$ or $H_0[bc_{11}=0]$, respectively, are assumed to be valid on grounds of the test results.

If each of the null hypotheses on the three interaction effects $\mathcal{AB}$, $\mathcal{AC}$, and $\mathcal{BC}$ has to be rejected, regardless of which one of the three confounded constants is deleted from the model, the conclusion is clearly that all two-factor interactions are significant and that the ranking process has reached the significant model. Naturally, there are many more possible results, all of which can not be discussed here, when testing the interaction effects under the condition of the confounding as contained in Example E. For example, the deleted constant may be ac_{21}, and $H_0[bc_{11}-bc_{21}=0]$ may be the only one of the three null hypotheses on interaction effects which does not have to be rejected. This also would mean that the significant model is reached in the ranking process.

Once the main effects in Example E become admissible for testing (i.e., once all interaction constants have been deleted from the model), their testing is straightforward since they are not affected by identities.

(Note. For a numerical illustration of Example E see Example 5 in Section 3.4.5.)

APPENDIX B

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) U. S. Naval Weapons Laboratory	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3 REPORT TITLE

THE METHOD AND USE OF NOVACOM, A PROGRAM FOR "NON-ORTHOGONAL" ANALYSIS OF VARIANCE AND COVARIANCE

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

5 AUTHOR(S) (First name, middle initial, last name)

Klaus Act

6 REPORT DATE 19 April 1968	7a. TOTAL NO. OF PAGES 20	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO b. PROJECT NO c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) TR-2108 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10 DISTRIBUTION STATEMENT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY

13 ABSTRACT

The report contains the description of a program ("NOVACOM") for the solution of problems in the area of analyzing data based on the general linear statistical model. While the detailed program documentation is given elsewhere, the present publication deals with the statistical method, the logical flow, and the use and application of NOVACOM in multiple linear regression and ("non-orthogonal") analysis of variance and covariance for crossed classifications with incomplete and unbalanced data. The method of NOVACOM is basically a backward ranking procedure applied to individual and/or groups of independent variables (concomitant) independent variables and/or ANOVA effects, respectively). The result of the ranking is a model ("significant model") which contains only significant concomitant independent variables and/or ANOVA effects. The method and use of the program is illustrated by examples of the statistical analysis of bodies of incomplete experimental data.

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED