IBM

Application Program

H20-0333-1

1130 Statistical System (1130-CA-06X)

User's Manual

The 1130 Statistical System performs four major statistical functions — regression analysis, factor analysis, analysis of variance, and polynomial fitting.

This manual contains, for each type of analysis performed, a description of the computational algorithms used, the form and content of the control cards, operating instructions, and sample problems.

Second Edition

H20-0333-1 is a major revision, obsoleting H20-0333-0 and incorporating
TNL N20-1029-0.

Specifications contained herein are subject to change from time to time. Any such change will be reported in subsequent revisions or Technical Newsletters.

Copies of this and other IBM publications can be obtained through IBM branch
offices. Address comments concerning the contents of this publication to
IBM, Technical Publications Department, 112 East Post Road, White Plains, N.Y. 10601

INTRODUCTION

The 1130 Statistical System contains four major analysis programs:

1. Stepwise Linear Regression
2. Factor Analysis
3. Analysis of Variance
4. Polynomial Fitting with Orthogonal Polynomials

Each of these analysis programs is composed of a number of subroutines, which are stored on the disk and are called into core storage when required for the execution of a particular job. The logic flow of the programs and the type of analysis to be performed is controlled by a main program, which reads the user-supplied parameter cards and calls in the appropriate link at the proper time.

Although the programs imply different techniques, a common approach can be used in executing a job for any analysis. Chapter 2 of this manual is divided into four parts that describe completely the necessary parameters and monitor control cards for execution.

Special features available with this package provide added user flexibility:

In all four major programs, data card formats can be specified by the user.

- Stepwise Linear Regression. Matrix input and output are allowed with a pooling option which provides for the combining of raw cross products matrices, either by addition or subtraction. This allows the combining of input from different sources, or of input that is available at different times, without requiring recalculation of these matrices. The subtraction feature gives flexibility for the handling of outlyers. Residuals are available on option.

- Factor Analysis. The pooling options, and matrix input/output, are available, as described above, for stepwise regression.

  Factor scores are calculated on option, and punched on option.

  Several options for the handling of communalities are available.

  Oblique and orthogonal rotations are allowed.

- Analysis of Variance. A table generation feature allows output from the factorial design analysis in standard format.

- Orthogonal Polynomials. These can be calculated for both equally spaced and unequally spaced intervals.

  Derivatives are calculated on option.

  Scaling of input is allowed on option.

CONTENTS

CHAPTER 1: GENERAL OPERATING INSTRUCTIONS

1.1 SYSTEM GENERATION

After preparing a disk with the monitor system (1130 Disk Monitor System, 1130-0S-001, described in manual C26-3750), the distributed source cards, preceded by a cold start card, can be loaded into the card reader hopper, and compiled and stored on the disk. It is not, however, necessary to store the total 1130 Statistical System on the disk before using any one of its four major programs. Within the discussion given for each specific program is information pertaining to loading the particular set of routines necessary for that analysis type.

The distributed decks consider that the operating system will use the 1132 Printer as output. If the console typewriter is to be used as the output device, monitor generation must consider this, and the IOCS card in each main program must be replaced by an IOCS card stating:

*IOCS (CARD, TYPEWRITER, DISK)

Identifying information for this exchange, which should take place before loading the statistical system, is listed below.

Program Name	IOCS Card Indentification (cc 73-80)
COREL	CORL 10
POLY	POLY 20
POL2	POL2 20
REGR	REGR 20
REGR2	RGR2 10
ANOVA	NOVA 20
ANOV2	NOV2 20
FCTR	FCTR 10
FCTR1	FCT1 20
FCTR2	FCT2 20
FCTR3	FCT3 20

In the routine PRNTB

a. Card PRNB 150 should be changed to read LIBF TYPEZ.

b. Cards PRNB 70 - PRNB 130 should be omitted.

In the decks distributed with this system, identifying labels are given in cc 73-76. These four characters do not allow labels to agree perfectly with names of programs. When referencing programs, keep this distinction in mind.

1.2 CONTROL CARDS

Each of the four programs included in the 1130 Statistical System requires monitor and program control cards. These cards are described in the job execution section of the specific program being considered. However, certain program control cards are standard for all programs. Their descriptions follow.

For any control card, numbers specified as integers (I) (this includes all numbers used for program control) should be specified as follows:

1. All numbers should be shifted to the right of their fields (right-justified) unless left justification is specifically called for.

2. Blanks and zeros are synonymous.

STANDARD PROGRAM CONTROL CARDS

Input/Output Units Card

The function of the input/output units card (Figure 1) is to assign logical unit numbers to each of the I/O devices used throughout the program. Each subroutine that requires the use of an I/O device has been programmed with symbolic unit designations. This card fixes a number to a specific I/O device.

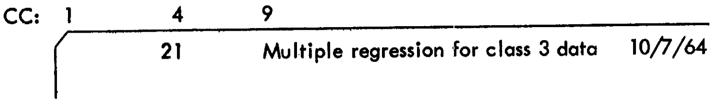
Column	Meaning
1-2	The unit for input of all control cards and source data. Normally, it is set equal to the logical number of the 1442 card reader, which is 02.
3-4	The unit used for card output of computed matrices. Normally, it is set equal to the logical number of the 1442 card punch, which is 02.
5-6	Output switch 0 - 1132 Printer output 1 - Typewriter output

```
CC:  1 2    3 4    5 6
    /0 2    0 2    0 0
    |
```

Figure 1. I/O units card (printer)

Job-Title Card

The job-title card (Figure 2) allows the user to assign a job number and title information for the job to be processed. This information is used only for labeling and is not used for processing in any program. The job number and title contained on the card are printed as the heading line on each page of output produced. The job number appears in the first four columns of any punched card output produced.



```
CC: 1      4      9
                  21     Multiple regression for class 3 data    10/7/64
```

Figure 2. Typical job-title card

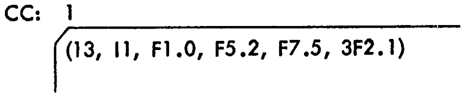
Column	Meaning
1-4	Job number
5-8	This field is not used.
9-80	Title information. These columns may contain any legitimate key-punchable characters that serve to identify the job.

Variable Format Card

Each program was designed to allow some flexibility in the input of data. Although 1130 FORTRAN does not allow the use of an object time format definition, a specially written format processing program is employed to enable the user to specify the format of his data by means of a FORTRAN-like statement. The format statement may contain almost all the specifications included in a normal FORTRAN format statement (as described in 1130 FORTRAN Language (C28-5933), pp. 11-15), with the following exceptions:

1. Only I, E, F, or X data specifications are allowed.
2. Continuation cards are not allowed.
3. The use of a slash (/) is not permitted.
4. Internal parentheses in the format specification are allowed.

The format card is punched with parentheses surrounding the specifications in columns 1-80, as shown in Figure 3.



```
CC: 1
   (I3, I1, F1.0, F5.2, F7.5, 3F2.1)
```

Figure 3. Format card example

Note: For each data card within an observation set (in case more than one card is required per observation), there must be a variable format card preceding the data deck. These format cards must be in the same order as are the cards in the observation sets. At most, the user will supply three format cards.

1.3 PROGRAM PAUSES AND MESSAGES

1. Pause 10. An illegal character in a numeric field has been encountered in reading data. The program will print the card and the approximate column where the error was detected. Pressing START on the card reader and console will cause the remaining data cards to be read and ignored. The next monitor control card, possibly signaling a new analysis, will be operated on. If this is not desired, the following should be done:

   It is possible that the format specification card is incorrect. If this is so, the entire deck to be analyzed must be rerun. However, if a specific data card is in error, the reader hopper and the stacker should be emptied. Pressing the nonprocess runout button will clear the card read-punch, and the second card in the stacker will be the card containing the error. After correcting this card, the user should place it and the third stacker card at the front of the deck that was withdrawn from the hopper, place this entire deck in the reader hopper, and press START on the card reader and console to continue processing.

2. Other error conditions are signaled by a printed message, and/or the program exits to the monitor. The monitor will read cards until a monitor control card is met (that is, the next job to be done), or will stop when the reader hopper is empty. For a list of the error messages, see chapter 5.

3. When an analysis is terminated successfully, an end-of-job message is printed, and control is relinquished to the monitor.

4. When the user calls for output on cards, a message is written reminding the operator to enter blank cards, if console entry switch 15 is not on. The computer then pauses to allow input of blank cards. If console entry switch 15 is on, no reminder is given. It is possible, in this case, to destroy the next analysis deck. See sections 1.4 and 1.5.

1.4 STACKING: SEQUENTIAL PROGRAM OPERATION

Stacking of jobs is permitted. Each job must be a complete deck, as defined in the job execution section of each program. However, when a program option card calls for output on the IBM 1442 Card Punch, the negative identification card following the input data must be succeeded by blank cards. For each matrix requested in factor analysis or regression analysis, it is wise to place at least $[n^2/5 + 2n + 2]$ blank cards behind the data deck, where n is the number of variables processed. For orthogonal polynomials, n + 1 blank cards should be included, where n is the order of the polynomial requested. When factor scores are to be punched, 2n blank cards should be included, where n is the number of observations processed.

It is advisable to place extra blank cards in the hopper, because an insufficient number could result in the destruction of a part of the next analysis deck. After an analysis is completed, cards are read until the next monitor control card is met.

1.5 TYPEWRITER AND PUNCHED CARD OUTPUT

For typewriter output, the 1130 Statistical System uses the same format statements as are used for the printer. A user electing to use this device heavily may desire to modify output using Assembler Language routines, calling on the typewriter tabulation feature.

All system programs, excepting analysis of variance, allow user selection of punched card output. This is discussed briefly in sections 1.3 (4) and 1.4. In the following, a detailed explanation of the mechanics of this operation is given.

Consider first that stacking is not being done; only one job is being run. If the user places one blank card behind his input deck, it is unnecessary to press the card reader and console start buttons to complete the reading of the input data. If the user has asked for punched output in his analysis definition (option card), an adequate number of blank cards should be in the hopper following the data (section 1.4). If this is not the case, the computer halts, waiting for the entry of blank cards. After these are placed in the reader hopper, the start button should be pressed on the card reader and console to continue processing.

If jobs have been stacked (section 1.4) and if, following the card with a negative identification field signifying end of data, there is another analysis deck (the next job), it is possible to destroy this next deck if punched output is being requested in the current job.

In this case, if console switch 15 is down (off), a message is written reminding the user to place blank cards in the hopper; then the computer pauses. If this occurs, the card reader hopper (which contains the next job to be run) should be emptied, the non process runout button on the card reader should be pressed, and the last two cards in the stacker (//XEQ and a LOCAL card) should be placed at the front of the next job to be run. Blank cards should then be placed in the reader hopper, followed by the next job to be run, and START pressed on the card reader and console to continue processing.

1.6 MACHINE AND SYSTEMS CONFIGURATION

The 1130 Statistical System is designed to operate on an 8K 1130 Computing System with disk storage (1131 Model II) and 1442 Card Read Punch; the 1132 Printer is optional. It is written to operate under the 1130 Disk Monitor System (1130-0S-001).

1.7 PROGRAMMING LANGUAGE

IBM 1130 FORTRAN and the IBM 1130 Assembler Language.

1.8 REFERENCE MATERIAL

IBM 1130 Disk Monitor System Reference Manual (C26-3750)
IBM 1130 Assembler Language (C26-5927)
IBM 1130 FORTRAN Language (C26-5933)

CHAPTER 2: PROGRAMS

2.1 STEPWISE LINEAR REGRESSION

From sets of observations numbering 499 or fewer, containing measures on a dependent variable y and n independent variables x_1, x_2, $\ldots x_n$, where the total number of variables is less than or equal to 30, the stepwise linear regression analysis will determine the coefficients of a linear equation of the form:

$$y = b_o + b_1x_1 + b_2x_2 + \ldots b_nx_n$$

which best approximates the observations in the least-squares sense.

The independent variables x_1, x_2, $\ldots$, x_n are entered into the equation on the basis of a variance criterion supplied by the user, which enables the program to determine which variable makes the greatest improvement in ''goodness of fit''. Similarly., variables are not entered, or removed, from the equation on the basis of a second variance criterion which indicates that the variable does not offer any significant improvement in the goodness of fit.

The general method of solution to determine the coefficients b_o, b_1, $\ldots b_m$ is to compute the matrix of correlation coefficients from the source data. This matrix will contain the correlations between all the independent variables and the dependent variable. By applying a Gaussian elimination inversion process, a stepwise inverse of the correlation matrix is computed. Multiplying this inverse by a vector containing the dependent variable correlated with each independent variable forms the normalized regression coefficients. The inversion process is carried out for one variable at a time. As each variable is processed, it is compared to the variance criterion to determine its significance. If the variable is to be entered, the coefficients for the equation containing a subset of the total number of variables in the analysis are computed and made available for printout and use in the next step of the analysis. Because of the nature of the computational process, the elements of several subsidiary statistics are also available. If the user elects to print each regression step as it is computed, these statistics will be printed with the regression coefficients.

The following book can be used as a reference: Ralston, A. and Wilf, H.S. Mathematical Methods for Digital Computers. New York: John Wiley and Sons, Inc., 1960.

2.1.1 Summary of Output Statistics

1. High and low value of each variable
2. Means of each variable
3. Standard deviation of each variable
4. Sample variance for each variable
5. Matrix of raw cross products

6. Matrix of residual cross products

7. Variance-covariance matrix

8. Matrix of correlation coefficients

9. Residual Standard Deviation

10. Standard error of the mean of the predicted dependent variable

11. Multiple correlation coefficient

12. Square of the multiple correlation coefficient = $\frac{\text{sum of squares due to regression}}{\text{adjusted total sum of squares}}$

13. Degrees of freedom

14. Regression coefficients - B

15. Standard error of regression coefficients

16. Partial correlation coefficients - $r_i = a_{in} / \sqrt{a_{ii}\ a_{nn}}$ where n denotes the dependent variable and a is an element of the stepwise inverse of the correlation matrix.

17. Normalized regression coefficients - B' ; $B_i' = B_i\ S_i\ /\ S_y$ where S_i is the standard deviation of the i^{th} independent variable

18. Standard error of normalized regression coefficients

19. For each data case, the predicted value and difference between the predicted value and the actual value

20. Analysis of variance table

2.1.2 Job Execution

To perform a regression analysis, the user must supply three sets of cards to the program:

1. Monitor control cards
2. Program control cards
3. Data cards

Descriptions of the form and content of each card set follow.

MONITOR CONTROL CARDS

The monitor control cards are necessary to initiate program loading from the disk and to establish the necessary communication with the monitor. A general description of cards may be found in IBM 1130 Disk Monitor System Reference Manual (C26-3750).

A regression analysis requires the following cards:

```
CC: 1   4   8       16-17
        ↓           ↙
    // XEQ REGR  03

    *LOCALREGR, FMTRD, PRNTB, DATRD, MXRAD, TRAN
    *LOCALCOREL, PRNT
    *LOCALREGR2, REGRE
```

The monitor control cards do not change from job to job within one analysis type, but must be included with every job processed. The first program operated on by this system should be preceded by a cold start card.

PROGRAM CONTROL CARDS

The program control cards communicate the data-specific parameters and output options to the program. There are five possible card types necessary for execution:

1. Input/output units card*
2. Job-title card*
3. Option card (described below)
4. Variable name card (described below)
5. Variable format card*

Four of the control cards are required in every job. The variable format card, which specifies data format, is not necessary if matrix data is to be processed.

OPTION CARD

Number of Variables (cc 1-2)

This field must be punched with a nonzero integer, n, which is less than or equal to 30. The value of integer n gives the total number of independent and dependent variables to be processed.

Input Type and Source (cc 3-4)

This field allows the user to specify the input device (1442 card reader or disk) and, indirectly, the type of input analysis to be undertaken in the input program. The three possible values that may be punched in this field are described below:

Value	Meaning
1	Raw data will be read from the 1442 card reader and transferred to the disk, where it will be retained until destroyed by input from

*See "General Operating Instructions", section 1.2.

Value	Meaning
1 (cont)	one of the four programs in this system. Until destroyed, option 2 (below) can be used to read this data from the disk.
2	Raw data will be read from the disk. Raw sums and raw sums of cross products will be accumulated. Data will be read until a negative number in the identification field is encountered (section 2. 1. 3).
3	A previously computed matrix, or matrices, will be read from the 1442 card reader. Matrix cards will be read until a negative job number field is encountered (see ''Pooling'', section 2. 1. 4).

Sequence Checking Within Observations (cc 5-6)

This field is used to indicate that raw data input from the card reader (cc 3-4 contains a 1) is to be sequence-checked. A value of zero or a blank field implies that no sequence check will be made. A value of one (1) implies that the cards will be sequence-checked. The sequence-checking process consists of an equal comparison check of the case identification field, for all cards in a case, and an ascending sequence check of the card number field. If an error in either of these conditions is encountered, the program prints a message, and the job is terminated.

Number of Variables on Card 1 (cc 7-8)

When a data vector contains more variables than will fit on one card, the user must indicate to the program the number of variables punched on each card. This field must be punched with the number of variables on the first card. If there is only one card per case, this field must be blank or zero.

Number of Variables on Card 2 (cc 9-10)

Same as cc 7-8, except that this field indicates the number of variables on the second card of the data.

Number of Variables on Card 3 (cc 11-12)

Same as cc 7-8, except that this field indicates the number of variables on the third card of the data.

Transformation Switch (cc 13-14)

If the value in this field is nonzero, a user-written transformation subroutine is called after each data record is read and before any computation takes place.

If the value in this field is zero or blank, the transformation subroutine is not called.

The use of transformations is discussed in section 2. 5. 1.

Output Raw Sums of Cross Products (cc 15-16)

This field is used to indicate whether the raw sums and sums of raw cross products matrix are to be printed, punched, printed and punched, or not presented.

The four (4) possible values of this field are described below. The computation to generate the matrix is performed even if the "no output" option is chosen.

Value	Meaning
0 or blank	No output.
1	Matrix will be printed.
2	Matrix will be printed and punched.
3	Matrix will be punched.

Punched output of the raw sums of cross products matrix includes the number of observations and the vector of raw sums and sums of squares. This entire output must be entered on the pooling option (section 2.1.3).

Output Residual Cross Products (cc 17-18)

This field is used to indicate whether the residual cross products matrix — defined as:

$$u_{ij} = c_{ij} - \frac{s_i s_j}{n} \qquad i, j = 1, 2 \ldots n$$

where c_{ij} are the elements of sums of raw cross products matrix, s_i, s_j are the raw sums of the $i^{\underline{th}}$ and $j^{\underline{th}}$ variables, respectively, and n is the number of cases — is to be printed, punched, printed and punched, or not presented.

The four (4) possible values are described above under "Output Raw Sums of Cross Products".

The matrix is computed even if the "no output" option is chosen.

Output Variance-Covariance Matrix (cc 19-20)

This field is used to indicate whether the variance-covariance matrix — defined as:

$$c_{ij} = \frac{u_{ij}}{n-1} \qquad i, j = 1, 2 \ldots n$$

where u_{ij} is an element of the residual cross products matrix, and n is the number of cases (or sum of weights) — is to be printed, printed and punched, punched, or not presented. The four (4) possible values that may occur in this field are as given for the above matrices.

There are no additional vectors or matrices punched with the punched output. The matrix is computed even if the ''no output'' option is chosen.

Output Correlation Matrix (cc 21-22)

This field is used to indicate whether the correlation matrix — defined by:

$$r_{ij} = \frac{c_{ij}}{s_i s_j} \qquad i, j = 1, 2 \ldots n$$

where c_{ij} is an element of the variance-covariance matrix, and s_i, s_j are the standard deviations of the $i^{\underline{th}}$ and $j^{\underline{th}}$ variables, respectively — is to be printed, punched, printed and punched, or not presented. The four (4) possible values contained in this field are given above under ''Output Raw Sums of Cross Products''.

The punched output of the correlation matrix includes the number of cases and cards containing the vectors of means and standard deviations.

The matrix is generated even if the ''no output'' option is chosen.

Compute and Print Predicted Values (and Residuals) (cc 23-24)

This field is used to indicate the regression step to begin computing and printing the predicted values, defined as:

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots b_k x_{ik}$$

where b_0, b_1, $\ldots b_k$ are the coefficients of the regression equation for the $k^{\underline{th}}$ step, and x_{ij} are the source data elements.

If this field is blank or zero, the predicted values are not computed. If the field is negative, predicted values are printed only on the final step.

If it contains a nonzero value, k, the predicted values are computed for each regression step equation containing k or more variables. For example, when k = 1, the predicted values for all step equations are printed. When k = 3, the predicted values for the equation containing three independent variables are printed. The printout also contains the actual value of y and the difference between the predicted and the actual values.

If predicted values are requested when equations with fewer than p variables are not desired, no predicted values are printed. That is, if column 23-24 contains a positive integer less than the integer in column 25-26, no predicted values are printed.

Print Steps of Regression (cc 25-26)

If this field contains a value of k, all step equations containing k or more independent variables are printed. For example, if all steps are desired, a value of one (1) forces the printout of the equations containing 1, 2, . . . m independent variables.

If this field is zero or blank, no printout occurs. Only a correlation matrix is calculated.

Pooling Option (cc 27-28)

When using the matrix input option (cc 3-4 are 03) and when pooling sums of squares and cross products (section 2. 1. 3), if the user desires that matrices be subtracted rather than added (of aid in deletion of outlyers), this field should be nonzero.

Number of the Dependent Variable (cc 29-30)

The regression analysis program uses the value punched in this field to rearrange the correlation matrix, means, standard deviation, and variable names vectors, such that the dependent variable always follows the independent variables. The user must therefore indicate to the program the number of the dependent variable in this field. The value punched must be greater than zero and less than or equal to 30. A value of zero implies a regression analysis is not desired, and the program will exit after the correlation analysis is complete.

Variance Criterion to Remove Variables (cc 31-34)

This field is used to determine whether an independent variable, when removed from the equation, significantly increases the sample residual variance. The significance of the increase is determined by comparing it to the variance criterion as punched in this field. If the computed variance measure is greater than the criterion, the variable is removed from the equation.

The form of the number to be punched is xxxx. A decimal point may replace any x. If there is no decimal point, the number is taken to be .xxxx. The size of this number depends on the information available to the particular analysis. If the user has no idea about the size to be used, a number between .005 and .05 may be acceptable.

It is possible for the user to set criterion levels for entry and removal of variables that cause cycling of variables into and out of the model. The program does not check this cycling possibility.

Variance Criterion to Enter Variables (cc 35-38)

This field has a similar function to the previous field, except that the value punched is to determine whether a variable is to enter the equation. A variable is entered into the equation if it significantly reduces the sample residual variance. The computed variance is compared to this criterion value to determine whether it does decrease this variance. The typical values and the form of the punched data are exactly as described above for the remove-variable criterion.

A number twice the size of the removal factor can be tried if the user is not sure of the correct number to be used in this field.

Tolerance For Ill Condition (cc 39-44)

A poorly conditioned matrix occurs when an independent variable is approximately a linear combination of other independent variables. This number is associated, in the program, with the size of the pivot element. If the pivot element is less than the tolerance level, the associated variable is not entered into the model on the iteration in which the condition occurs. A tolerance level of zero is not to be advised, unless the user is sure that his matrix is not ill-conditioned.

Regression Analysis Option Card Summary

Column	Meaning	
1-2	Number of variables	
3-4	Input type and source 1 - Raw data input from card reader 2 - Raw data input from disk 3 - Matrix input from card reader	
5-6	*Check sequence of raw data input 0 - No 1 - Yes	
7-8	*Number of variables on card 1	(must be blank if there is only one card per observation)
9-10	*Number of variables on card 2	
11-12	*Number of variables on card 3	
13-14	*Transformation switch 0 - No transformation 1 - Transformation	
15-16	**Output raw cross products matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch	
17-18	**Output adjusted cross products matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch	
19-20	**Output variance-covariance matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch	

* Not pertinent when matrix input is used.

** When correlation matrix input is used, matrices are not available for output.

Column	Meaning
21-22	**Output correlation matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch
23-24	*Output predicted values 1 - Print predicted values for last step only. 0 - Do not print predicted values. k - Print predicted values for models containing k or more independent variables
25-26	Output steps of regression 0 - Print no regression steps. Exit after correlation analysis. k - Print all steps for models containing k or more independent varia
27-28	Pooling option (see sections 2.5.3 and 2.1.4) Zero - Add matrices with ID = 1 Nonzero - Subtract matrices with ID = 1
29-30	Number of dependent variable
31-34	Variance criterion to remove variables
35-38	Variance criterion to enter variables
39-44	Tolerance for colinearity

VARIABLE NAME CARD (Figure 4)

In the multiple regression program there are a number of matrix printouts that the user may request. The variables in the matrix may be assigned a four-character name to aid in the identification of the output. The card is punched in four-column fields, and each field corresponds to the variable to be identified (for example, field 3 (columns 9-12) will be the name of row and column 3 on all matrix output). At most, 20 names can appear on one card. If there are more than 20 variables in the analysis, a second card having the same format as the first must be included in the control card deck.

Column	Meaning
1-4	Name of variable 1.
5-8	Name of variable 2.
.....	....
(4N-3) - (4N)	Name of variable N.

* Not pertinent when matrix input is used.

** When correlation matrix input is used, matrices are not available for output.

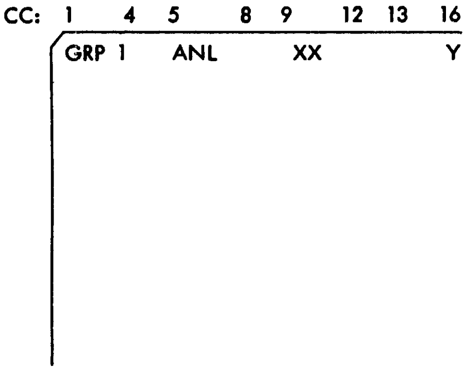


Figure 4. Variable name card example

2.1.3 Data Input

Raw data input to the regression program consists of a set of observations made on several different variables. The variables for each observation are punched on one, two, or three cards, according to the following general form:

Field	Type	Meaning
1	Integer (I)	Identification field. Any numeric information that serves to identify the particular observation is punched in this field. It must be greater than zero, and should be different for each observation.
2	Integer (I)	Card number within observation. If it is not possible for one card to contain all the variables, they may be continued on a second and a third card, as necessary. The user has the option of sequence checking the cards to ensure that all cards within a case are together, and that the order of cards is consistent. If the option is chosen (cc 5-6 on option card), this field must be punched with an integer that is in ascending sequence for all cards in the case. If sequence checking is not desired, the field may be blank and may consist of one blank column.
3,4,...,n etc.	Floating point (F)	Observation on variable x_1. Any number may be punched in this field. Decimal points are not required. The remaining fields on the card are reserved for variable observations. If there are more variables than can fit on the first card, a second and a third card may be used.

The particular card columns for each field are arbitrary.

Following the data deck, the user must include a card containing a negative integer in the identification field. This card signals the end of data.

2.1.4 Matrix Input/Output

It is possible to obtain punched card output of a number of matrices (see section 2.5.3) and vectors with the regression program. This program is designed to also input some of these matrices, at a later time, for further analysis or processing. In addition, matrices from another program or source, if punched in the proper format, may also be used as input.

This section is devoted to a description of various possible forms of analysis with the output options available in each program.

Format Description

Matrices are punched rowwise, five elements to a card, in the FORTRAN E or floating-point format. Each card is identified as to its job number, matrix number, row number, and column number of the first element on the card. The specific card columns occupied by the identification fields, and matrix elements, are shown below:

Column	Meaning
1-4	Job number
5-6	Matrix identification number (section 2.5.3)
7-8	Column number of first element on card
9-10	Row number
11-24	Matrix element (±0.XXXXXXXE±NN)
25-38	Matrix element
39-52	Matrix element
53-66	Matrix element
67-80	Matrix element

Note that all matrices are punched and read under a fixed format. Hence, a variable format card is not allowed when using punched card matrices as input.

Most matrices have a unique identification number. However, there are a few cases where two or three vectors have the same identification and are always punched together. For these cases, see section 2.5.3.

Regression with Correlation Matrix Input

The punched output option of the correlation matrix includes the punchout of the number of cases (matrix 21), and means and standard deviation vectors (matrix 23). This complete output can be used as input to initiate another analysis without the necessity of reprocessing the raw data used to generate the matrices.

To use the correlation matrix set as input, the user places the punched output behind the variable names card, followed by a card that contains a negative number in the job number field. The program reads the number of cases, means, standard deviations, and correlation matrix, storing each in its appropriate location, and then initiates the analysis as specified on the option card. No matrix output is possible in this case. Also, observations are not read; hence, predicted values and certain summary statistics are not available.

Pooling Sums of Squares and Cross Products (cc 27-28)

In the regression and factor analysis programs, a considerable amount of processing time is devoted to accumulating raw sums and raw cross products as each data vector is read. If there is a large amount of source data, or if there is some logical division in the data set, it is frequently desirable to obtain partial punched output of the raw sums of squares and cross products. These partial outputs can then be added together to complete the total analysis in another job.

Both programs allow this type of analysis. By choosing the punchout option for this matrix, the program includes in the punchout the number of cases, and raw sums and sums of squares vectors, in addition to the raw cross products matrix. Any number of these matrices may be punched and used later to complete the analysis. The user simply stacks each output set, one after the other, following the variable names card. The program reads the matrices, examining the matrix identification and row and column numbers to determine the location or group of locations to which the matrix is to be added. The read-add operation is terminated when a card with a blank or negative job number field is encountered, unless the pooling option (cc 27-28 of the option card) is nonzero. In this case, the read-add operation terminates at the first detection of a blank or negative job number field, and the second (succeeding) matrix is subtracted from the previous matrix. This operation is terminated when the second blank or negative job number is encountered.

If the subtraction option is used, the second set of matrices must also include its associated raw sums and sums of squares vectors for proper analysis completion.

Predicted values are not available with this option. Also, high and low values are not calculated for the observations on variables.

If outlyers are detected, the user has two options available if he wishes to reanalyze ignoring these outlyers:

1. He can eliminate the data cards containing the outlyers, and rerun the entire analysis.

2. He can prepare cards according to the format given above under "Format Description", either by hand or by using the program. To use the program, he must run the analysis using only data cards associated with outlyers. The option card must request raw cross products matrix output, and may note that the dependent variable is zero, so that the analysis will terminate after the correlation matrix is calculated.
   If the user allows an entire regression analysis to be computed using only the outlyer cards, a termination with some error condition may result — for example, mean square nonpositive. In any case, matrices 1, 21, and 22 will be punched (see section 2.5.3). These matrices should be used as the second set of matrices for input using the subtraction option.

2.1.5 Operating Instructions

A. Using the regression analysis program when the total 1130 Statistical System has not been stored on the disk

If the user wishes to load only the set of programs that allow regression analyses, the following programs must be compiled or assembled and stored on the disk. Each deck begins with a card punched as

```
//FOR
```

and ends with an

```
*STORE
```

card.

The user should use a disk containing the 1130 Disk Monitor System, as described in section 1.1. The following decks should be preceded by a cold start card, placed in the card reader hopper, and the buttons IMMEDIATE STOP (console), RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. A blank card should be placed after the last deck in the card reader hopper.

DECKS-LABELS: REGR-REGR; **COREL-CORL; **PRNT-PRNT; *FMTRD-FMRD; *DATRD-DTRD; *PRNTB-PRNB; *GMPYX-GMPY; *GDIVX-GDIV; **MXRAD-MXRD; REGR2-RGR2; REGRE-RGRE; *FMAT-FMAT; TRAN-TRAN.

In addition, regression and factor analysis programs must reside on the disk together; section 2.2.7 names additional routines to be placed on the disk.

B. Execution from disk

Once the component subroutines and main calling programs are on the disk, the execution of a job requires the monitor control cards, program control cards, and data cards to be

*Used in all four analysis types
**Used in factor analysis

placed in the card reader. The deck should be preceded by a cold start card. To initiate processing, the buttons IMMEDIATE STOP and RESET (console), START (card reader and printer), and PROGRAM LOAD (console) should be pressed. The order in which the cards are placed in the card reader for either matrix or raw data input is shown in Figures 5, 6, and 7.

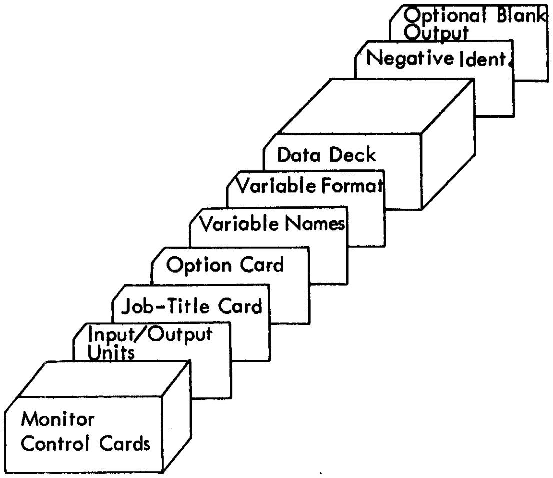


Figure 5. Regression card order — card reader input

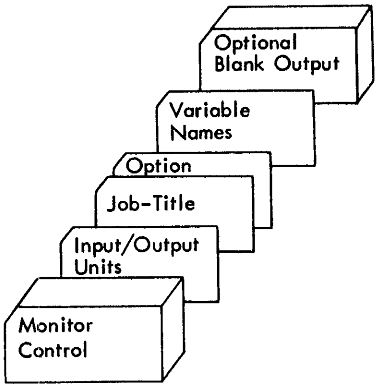


Figure 6. Regression card order — disk input

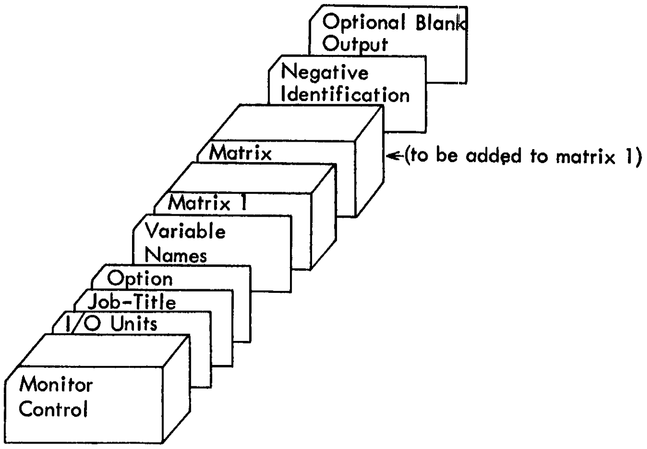


Figure 7. Regression card order — matrix input

2.1.6 Sample Problem

INPUT

```
// XEQ REGR     03
*LOCALREGR,FMTRD,PRNTB,DATRD,MXRAD,TRAN
*LOCALREGR2,REGRE
*LOCALCOREL,PRNT
020200
2222     STEPWISE TEST ONE
06010000      0001010102-1010006.500.300.00010
  P1  P2  P3  P4  P5  P6
(2I2,1X  ,F5.2,F5.0,2F5.2,2F5.0)
0101 002500002502500001500003400064
0201 013000002102100000870003600065
0301 003500002202200000430004100082
0401 001750000900130001800001500023
0501 003000002302300002000003300064
0601 002000001000060003300001300016
0701 005500000700140003400001600012
0801 006000000600080005000001100027
0901 001300000800270001500001900048
1001 005000001800360001800002700050
1101 005000000300100001400001400012
1201 003000000800270001000002500013
1301 002000000600300001500002100020
1401 002000000800100002500001800023
1501 001000002202200001100004600118
1601 004000001301300002800001700050
1701 000500002600120000730004800063
1801 000250002302300000100003600150
1901 014000000300100003500000500072
2001 002500001500250000280003300054
2101 003500002801400000010004600109
2201 003500000600060005000001000010
2301 002500003503500005700003800125
2401 000500001100200003400001600044
2501 002000001101100000500002000048
2601 007000003203200006600003800105
2701 004000000800100004500001200009
2801 015000002302300000150004900130
2901 001000003803800002200004300160
3001 003500001500500001500003300048
```

```
3101 0130000006001200037000000900036
3201 0020000025025000010000003500150
3301 0120000005001700003000002100078
3401 0040000009000750019000001700023
3501 0030000007003500026000001200042
3601 0080000002002000000220000300072
3701 0090000006000860025000001500020
3801 0060000012004000012000002000036
3901 0080000026001600011000003500056
4001 0015000015003000016000002900036
4101 0070000010000900100000001200026
4201 0080000028028000004200004000108
4301 0020000034034000000900004200106
4401 0060000004000800036000001100016
4501 0150000032032000018000044000104
4601 0170000011011000023000001400047
4701 0160000002000500018000001100027
4801 0030000018001600011000003200012
4901 0060000003000400013000001500007
5001 0140000008001100020000001700018
5101 0060000014000900007000002900028
5201 0018000012002400015000002100025
5301 0150000003001500008000001300011
5401 0180000006005500057000000900020
5501 0050000012002000041000001600014
5601 0300000011011000020000002200038
5701 0290000008008000010000002200103
5801 0018000024024000011000003800106
5901 0130000026026000017000003800063
6001 0190000029029000480000002900208
6101 0110000017017000016000002500032
6201 0100000015005000035000001900028
6301 0060000010005000010000002600032
6401 0050000022022000012000003900100
6501 0010000015005000008000002900050
6601 0170000009003000130000001000080
6701 0050000030035000009000005800065
6801 0013000010001300090000001000025
-1
```

PUNCHED CORRELATION MATRIX OUTPUT

```
2222 4 1 1 0.1000000E 01-0.1764650E 00 0.5134508E-02 0.2554817E 00-0.1956896E 00
2222 4 1 2-0.1764650E 00 0.1000000E 01 0.8679906E 00 0.1007193E 00 0.8791197E 00
2222 4 1 3 0.5134508E-02 0.8679906E 00 0.1000000E 01 0.1258658E 00 0.7519358E 00
2222 4 1 4 0.2554817E 00 0.1007193E 00 0.1258658E 00 0.1000000E 01-0.1404377E 00
2222 4 1 5-0.1956896E 00 0.8791197E 00 0.7519358E 00-0.1404377E 00 0.1000000E 01
2222 4 1 6 0.8205512E-01 0.7496167E 00 0.7860574E 00 0.3425673E 00 0.6451673E 00
2222 4 6 1 0.8205512E-01
2222 4 6 2 0.7496167E 00
2222 4 6 3 0.7860574E 00
2222 4 6 4 0.3425673E 00
2222 4 6 5 0.6451673E 00
2222 4 6 6 0.1000000E 01
222223 1 1 0.6995588E 01 0.6473750E 01
222223 1 2 0.1525000E 02 0.9357534E 01
222223 1 3 0.1042514E 02 0.1162702E 02
222223 1 4 0.3099554E 01 0.5997127E 01
222223 1 5 0.2539706E 02 0.1247940E 02
222223 1 6 0.5679412E 02 0.4355013E 02
222221 1 1 0.6800001E 02
```

OUTPUT

```
// XEQ REGR     03
*LOCALREGR,FMTRD,PRNTB,DATRD,MXRAD,TRAN
*LOCALREGR2,REGRE
*LOCALCOREL,PRNT
```

```
        STEPWISE TEST ONE                                                        JOB   2222     PAGE     0

         NUMBER OF VARIABLES                  6
         INPUT TYPE                           1
         SEQUENCE CHECK                       0
         VARIABLES ON CARD 1                  0
         VARIABLES ON CARD 2                  0
         VARIABLES ON CARD 3                  0
         TRANSFORMATION SWITCH                0
         OUTPUT RAW CROSS PRODUCTS            1
         OUTPUT RESIDUAL CROSS PRODUCTS       1
         PRINT PREDICTED VALUES              -1
         PRINT STEPS                          1
         POOLING OPTION                       0
         DEPENDENT VARIABLE                   6
         F-LEVEL TO REMOVE VARIABLES      0.500
         F-LEVEL TO ENTER VARIABLES       0.300
         TOLERANCE VALUE                0.00010
         OUTPUT VARIANCE - COVARIANCE         1
         OUTPUT CORRELATION                   2
(2I2,1X  ,F5.2,F5.0,2F5.2,2F5.0)
```

```
        STEPWISE TEST ONE                                                        JOB   2222     PAGE     1

                                        MATRIX OF RAW CROSS-PRODUCTS


VARIABLE        P1              P2              P3              P4              P5              P6
    P1     0.61357E 04     0.65381E 04     0.49851E 04     0.21390E 04     0.11022E 05     0.28566E 05
    P2     0.65381E 04     0.21681E 05     0.17138E 05     0.35929E 04     0.33215E 05     0.79363E 05
    P3     0.49851E 04     0.17138E 05     0.16448E 05     0.27853E 04     0.25314E 05     0.66929E 05
    P4     0.21390E 04     0.35929E 04     0.27853E 04     0.30629E 04     0.46487E 04     0.17964E 05
    P5     0.11022E 05     0.33215E 05     0.25314E 05     0.46487E 04     0.54295E 05     0.12157E 06
    P6     0.28566E 05     0.79363E 05     0.66929E 05     0.17964E 05     0.12157E 06     0.34641E 06
```

```
        STEPWISE TEST ONE                                                        JOB   2222     PAGE     2

                                      MATRIX OF RESIDUAL CROSS-PRODUCTS


VARIABLE        P1              P2              P3              P4              P5              P6
    P1     0.28079E 04    -0.71622E 03     0.25893E 02     0.66455E 03    -0.10592E 04     0.15499E 04
    P2    -0.71622E 03     0.58667E 04     0.63273E 04     0.37869E 03     0.68782E 04     0.20467E 05
    P3     0.25893E 02     0.63273E 04     0.90575E 04     0.58802E 03     0.73100E 04     0.26667E 05
    P4     0.66455E 03     0.37869E 03     0.58802E 03     0.24096E 04    -0.70419E 03     0.59945E 04
    P5    -0.10592E 04     0.68782E 04     0.73100E 04    -0.70419E 03     0.10434E 05     0.23492E 05
    P6     0.15499E 04     0.20467E 05     0.26667E 05     0.59945E 04     0.23492E 05     0.12707E 06
```

```
        STEPWISE TEST ONE                                                          JOB    2222      PAGE     3
                                              VARIANCE - COVARIANCE MATRIX


VARIABLE        P1          P2          P3          P4          P5          P6
    P1     0.41909E 02 -0.10689E 02  0.38647E 00  0.99187E 01 -0.15809E 02  0.23134E 02
    P2    -0.10689E 02  0.87563E 02  0.94437E 02  0.56521E 01  0.10266E 03  0.30548E 03
    P3     0.38647E 00  0.94437E 02  0.13518E 03  0.87764E 01  0.10910F 03  0.39802E 03
    P4     0.99187E 01  0.56521E 01  0.87764E 01  0.35965E 02 -0.10510F 02  0.89470E 02
    P5    -0.15809E 02  0.10266E 03  0.10910E 03 -0.10510E 02  0.15573E 03  0.35063E 03
    P6     0.23134E 02  0.30548E 03  0.39802E 03  0.89470E 02  0.35063E 03  0.18966E 04
```

```
        STEPWISE TEST ONE                                                          JOB    2222      PAGE     4

         SUMMARY STATISTICS          NO.OF CASES=    68

             VARIABLE                    LOW         HIGH          AVERAGE       STD. DEV.      VARIANCE


              1    P1          0.25000E 00   0.30000E 02   0.69955E 01   0.64737E 01   0.41909E 02
              2    P2          0.20000E 01   0.38000E 02   0.15250E 02   0.93575E 01   0.87563E 02
              3    P3          0.40000E 00   0.38000E 02   0.10425E 02   0.11627E 02   0.13518E 03
              4    P4          0.10000E-01   0.48000E 02   0.30995E 01   0.59971E 01   0.35965E 02
              5    P5          0.50000E 01   0.58000E 02   0.25397E 02   0.12479E 02   0.15573E 03
              6    P6          0.70000E 01   0.20800E 03   0.56794E 02   0.43550E 02   0.18966E 04

READY THE PUNCH WITH BLANK CARDS AND PRESS START ON THE PUNCH AND CONSOLE.  TURN CONSOLE SWITCH 15 ON.
```

```
        STEPWISE TEST ONE                                                          JOB    2222      PAGE     5
                                          MATRIX OF CORRELATION COEFFICIENTS


VARIABLE        P1          P2          P3          P4          P5          P6
    P1     0.10000E 01 -0.17646E 00  0.51345E-02  0.25548E 00 -0.19568E 00  0.82055E-01
    P2    -0.17646E 00  0.10000E 01  0.86799E 00  0.10071E 00  0.87911E 00  0.74961E 00
    P3     0.51345E-02  0.86799E 00  0.10000E 01  0.12586E 00  0.75193E 00  0.78605E 00
    P4     0.25548E 00  0.10071E 00  0.12586E 00  0.10000E 01 -0.14043E 00  0.34256E 00
    P5    -0.19568E 00  0.87911E 00  0.75193E 00 -0.14043E 00  0.10000E 01  0.64516E 00
    P6     0.82055E-01  0.74961E 00  0.78605E 00  0.34256E 00  0.64516E 00  0.10000E 01
```

```
        STEPWISE TEST ONE                                                                   JOB    2222      PAGE      6

     REGRESSION ANALYSIS

     DEPENDENT VARIABLE                P6
     RESIDUAL STANDARD DEVIATION  27.1238
     STANDARD ERROR OF THE MEAN    3.2892
     MULTIPLE R                    0.7860
     MULTIPLE RSQR                 0.6178


    VARIABLE ENTERED                   P3


   VARIABLE            B - COEF   STD ERROR OF B          PARTIAL-R         BETA-COEF    STD ERROR OF BETA

       P3                2.9442        0.2850              0.7860            0.7860            0.0760


  CONSTANT             26.0998


                               ANALYSIS OF VARIANCE TABLE

             SOURCE               D.F.      SUM OF SQUARES    MEAN SQUARE        F

             MEAN                   1       0.21933E 06       0.21933E 06
             REGRESSION             1       0.78516E 05       0.78516E 05    0.10672E 03
             ERROR                 66       0.48556E 05       0.73570E 03
```

```
        STEPWISE TEST ONE

     REGRESSION ANALYSIS

     DEPENDENT VARIABLE                P6
     RESIDUAL STANDARD DEVIATION  25.0821
     STANDARD ERROR OF THE MEAN    3.0416
     MULTIPLE R                    0.8235
     MULTIPLE RSQR                 0.6781


    VARIABLE ENTERED                   P4


   VARIABLE            B - COEF   STD ERROR OF B          PARTIAL-R         BETA-COEF    STD ERROR OF BETA

       P3                2.8275        0.2656              0.7971            0.7548            0.0709
       P4                1.7976        0.5150              0.3972            0.2475            0.0709


  CONSTANT             21.7445


                               ANALYSIS OF VARIANCE TABLE

             SOURCE               D.F.      SUM OF SQUARES    MEAN SQUARE        F

             MEAN                   1       0.21933E 06       0.21933E 06
             REGRESSION             2       0.86180E 05       0.43090E 05    0.68493E 02
             ERROR                 65       0.40892E 05       0.62911E 03
     STEPWISE TEST ONE                                                                      JOB    2222      PAGE      8

  REGRESSION ANALYSIS

  DEPENDENT VARIABLE                     P6
  RESIDUAL STANDARD DEVIATION       23.9328
  STANDARD ERROR OF THE MEAN         2.9022
  MULTIPLE R                         0.8435
  MULTIPLE RSQR                      0.7115


 VARIABLE ENTERED                  P5


VARIABLE          B - COEF    STD ERROR OF B          PARTIAL-R         BETA-COEF     STD ERROR OF BETA

   P3              1.9583        0.4079                 0.5145            0.5228              0.1089
   P4              2.3123        0.5266                 0.4811            0.3184              0.0725
   P5              1.0355        0.3808                 0.3217            0.2967              0.1091


CONSTANT          2.9105


                              ANALYSIS OF VARIANCE TABLE

                 SOURCE            D.F.      SUM OF SQUARES    MEAN SQUARE         F

                 MEAN               1        0.21933E 06      0.21933E 06
                 REGRESSION         3        0.90415E 05      0.30138E 05     0.52617E 02
                 ERROR             64        0.36658E 05      0.57278E 03
```

```
     STEPWISE TEST ONE                                                                      JOB    2222      PAGE      9

  REGRESSION ANALYSIS

  DEPENDENT VARIABLE                     P6
  RESIDUAL STANDARD DEVIATION       23.9726
  STANDARD ERROR OF THE MEAN         2.9071
  MULTIPLE R                         0.8456
  MULTIPLE RSQR                      0.7150


 VARIABLE ENTERED                  P1


VARIABLE          B - COEF    STD ERROR OF B          PARTIAL-R         BETA-COEF     STD ERROR OF BETA

   P1              0.4272        0.4813                 0.1111            0.0635              0.0715
   P3              1.8967        0.4145                 0.4994            0.5063              0.1106
   P4              2.2333        0.5349                 0.4654            0.3075              0.0736
   P5              1.1167        0.3923                 0.3375            0.3200              0.1124


CONSTANT         -1.2530


                              ANALYSIS OF VARIANCE TABLE

                 SOURCE            D.F.      SUM OF SQUARES    MEAN SQUARE         F

                 MEAN               1        0.21933E 06      0.21933E 06
                 REGRESSION         4        0.90867E 05      0.22716E 05     0.39529E 02
                 ERROR             63        0.36205E 05      0.57468E 03
```

```
STEPWISE TEST ONE                                                      JOB   2222     PAGE    10


                          PREDICTED VALUES

CASE              ACTUAL              PREDICTED                RESIDUAL


   1           0.6400E 02           0.8855E 02           -0.2455E 02
   2           0.6500E 02           0.8627E 02           -0.2127E 02
   3           0.8200E 02           0.8871E 02           -0.6717E 01
   4           0.2300E 02           0.2273E 02            0.2682E 00
   5           0.6400E 02           0.8497E 02           -0.2097E 02
   6           0.1600E 02           0.2262E 02           -0.6627E 01
   7           0.1200E 02           0.2921E 02           -0.1721E 02
   8           0.2700E 02           0.2627E 02            0.7213E 00
   9           0.4800E 02           0.2899E 02            0.1900E 02
  10           0.5000E 02           0.4188E 02            0.8116E 01
  11           0.1200E 02           0.2154E 02           -0.9541E 01
  12           0.1300E 02           0.3530E 02           -0.2230E 02
  13           0.2000E 02           0.3209E 02           -0.1209E 02
  14           0.2300E 02           0.2718E 02           -0.4183E 01
  15           0.1180E 03           0.9473E 02            0.2326E 02
  16           0.5000E 02           0.5035E 02           -0.3517E 00
  17           0.6300E 02           0.5647E 02            0.6528E 01
  18           0.1500E 03           0.8290E 02            0.6709E 02
  19           0.7200E 02           0.2002E 02            0.5197E 02
  20           0.5400E 02           0.4203E 02            0.1196E 02
  21           0.1090E 03           0.7818E 02            0.3081E 02
  22           0.1000E 02           0.2371E 02           -0.1371E 02
  23           0.1250E 03           0.1213E 03            0.3631E 01
  24           0.4400E 02           0.2821E 02            0.1578E 02
  25           0.4800E 02           0.4391E 02            0.4082E 01
  26           0.1050E 03           0.1196E 03           -0.1461E 02
  27           0.9000E 01           0.2580E 02           -0.1680E 02
  28           0.1300E 03           0.1038E 03            0.2616E 02
  29           0.1600E 03           0.1241E 03            0.3581E 02
  30           0.4800E 02           0.4992E 02           -0.1928E 01
  31           0.3600E 02           0.2489E 02            0.1110E 02
  32           0.1500E 03           0.8833E 02            0.6166E 02
  33           0.7800E 02           0.3121E 02            0.4678E 02
  34           0.2300E 02           0.2510E 02           -0.2106E 01
  35           0.4200E 02           0.2587E 02            0.1612E 02
  36           0.7200E 02           0.7851E 02           -0.6515E 01
  37           0.2000E 02           0.2655E 02           -0.6557E 01
  38           0.3600E 02           0.3391E 02            0.2087E 01
  39           0.5600E 02           0.4674E 02            0.9257E 01
  40           0.3600E 02           0.4103E 02           -0.5037E 01
```

```
STEPWISE TEST ONE                                                      JOB   2222     PAGE    11


                          PREDICTED VALUES

CASE              ACTUAL              PREDICTED                RESIDUAL


  41           0.2600E 02           0.3917E 02           -0.1317E 02
  42           0.1080E 03           0.1093E 03           -0.1323E 01
  43           0.1060E 03           0.1130E 03           -0.7004E 01
  44           0.1600E 02           0.2315E 02           -0.7151E 01
  45           0.1040E 03           0.1190E 03           -0.1500E 02
  46           0.4700E 02           0.4764E 02           -0.6450E 00
  47           0.2700E 02           0.2283E 02            0.4164E 01
  48           0.1200E 02           0.4125E 02           -0.2925E 02
  49           0.7000E 01           0.2172E 02           -0.1472E 02
  50           0.1800E 02           0.3026E 02           -0.1226E 02
  51           0.2800E 02           0.3696E 02           -0.8966E 01
  52           0.2500E 02           0.3087E 02           -0.5869E 01
  53           0.1100E 02           0.2430E 02           -0.1330E 02
  54           0.2000E 02           0.3964E 02           -0.1964E 02
  55           0.1400E 02           0.3170E 02           -0.1770E 02
  56           0.3800E 02           0.6146E 02           -0.2346E 02
  57           0.1030E 03           0.5311E 02            0.4988E 02
  58           0.1060E 03           0.8993E 02            0.1606E 02
  59           0.6300E 02           0.9984E 02           -0.3684E 02
  60           0.2080E 03           0.2014E 03            0.6543E 01
  61           0.3200E 02           0.6718E 02           -0.3518E 02
  62           0.2800E 02           0.4153E 02           -0.1353E 02
  63           0.3200E 02           0.4206E 02           -0.1006E 02
  64           0.1000E 03           0.8884E 02            0.1115E 02
  65           0.5000E 02           0.4283E 02            0.7169E 01
  66           0.8000E 02           0.5190E 02            0.2809E 02
  67           0.6500E 02           0.1340E 03           -0.6905E 02
  68           0.2500E 02           0.3303E 02           -0.8035E 01

JOB COMPLETED
```

CORRELATION MATRIX INPUT

```
// XEQ REGR     03
*LOCALREGR,FMTRD,PRNTB,DATRD,MXRAD,TRAN
*LOCALREGR2,REGRE
*LOCALCOREL,PRNT
020200
2222    STEPWISE TEST ONE
06030000     000000000000010006.500.300.00010
  P1  P2  P3  P4  P5  P6
2222 4 1 1 0.1000000E 01-0.1764650E 00 0.5134508E-02 0.2554817E 00-0.1956896E 00
2222 4 1 2-0.1764650E 00 0.1000000E 01 0.8679906E 00 0.1007193E 00 0.8791197E 00
2222 4 1 3 0.5134508E-02 0.8679906E 00 0.1000000E 01 0.1258658E 00 0.7519358E 00
2222 4 1 4 0.2554817E 00 0.1007193E 00 0.1258658E 00 0.1000000E 01-0.1404377E 00
2222 4 1 5-0.1956896E 00 0.8791197E 00 0.7519358E 00-0.1404377E 00 0.1000000E 01
2222 4 1 6 0.8205512E-01 0.7496167E 00 0.7860574E 00 0.3425673E 00 0.6451673E 00
2222 4 6 1 0.8205512E-01
2222 4 6 2 0.7496167E 00
2222 4 6 3 0.7860574E 00
2222 4 6 4 0.3425673E 00
2222 4 6 5 0.6451673E 00
2222 4 6 6 0.1000000E 01
222223 1 1 0.6995588E 01 0.6473750E 01
222223 1 2 0.1525000E 02 0.9357534E 01
222223 1 3 0.1042514E 02 0.1162702E 02
222223 1 4 0.3099554E 01 0.5997127E 01
222223 1 5 0.2539706E 02 0.1247940E 02
222223 1 6 0.5679412E 02 0.4355013E 02
222221 1 1 0.6800001E 02
  -1
```

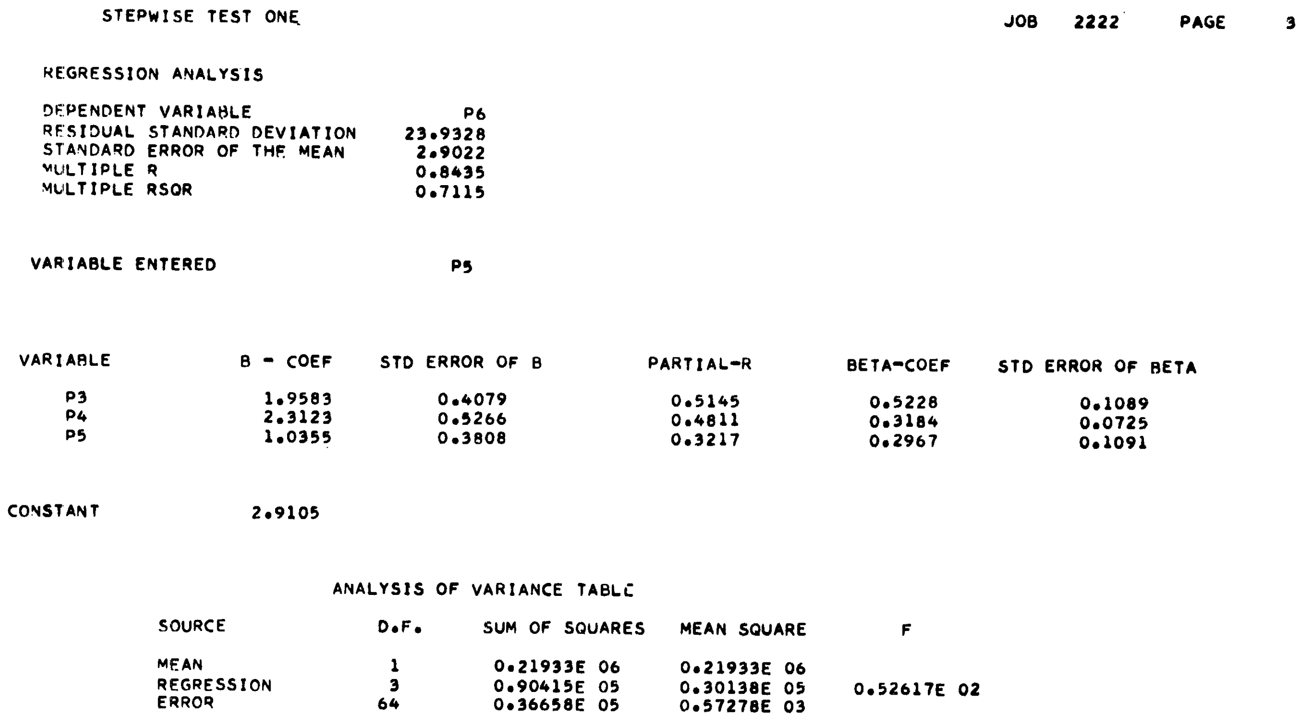
OUTPUT

```
// XEQ REGR     03
*LOCALREGR•FMTRD•PRNTB•DATRD•MXRAD•TRAN
*LOCALREGR2•REGRE
*LOCALCOREL•PRNT
```

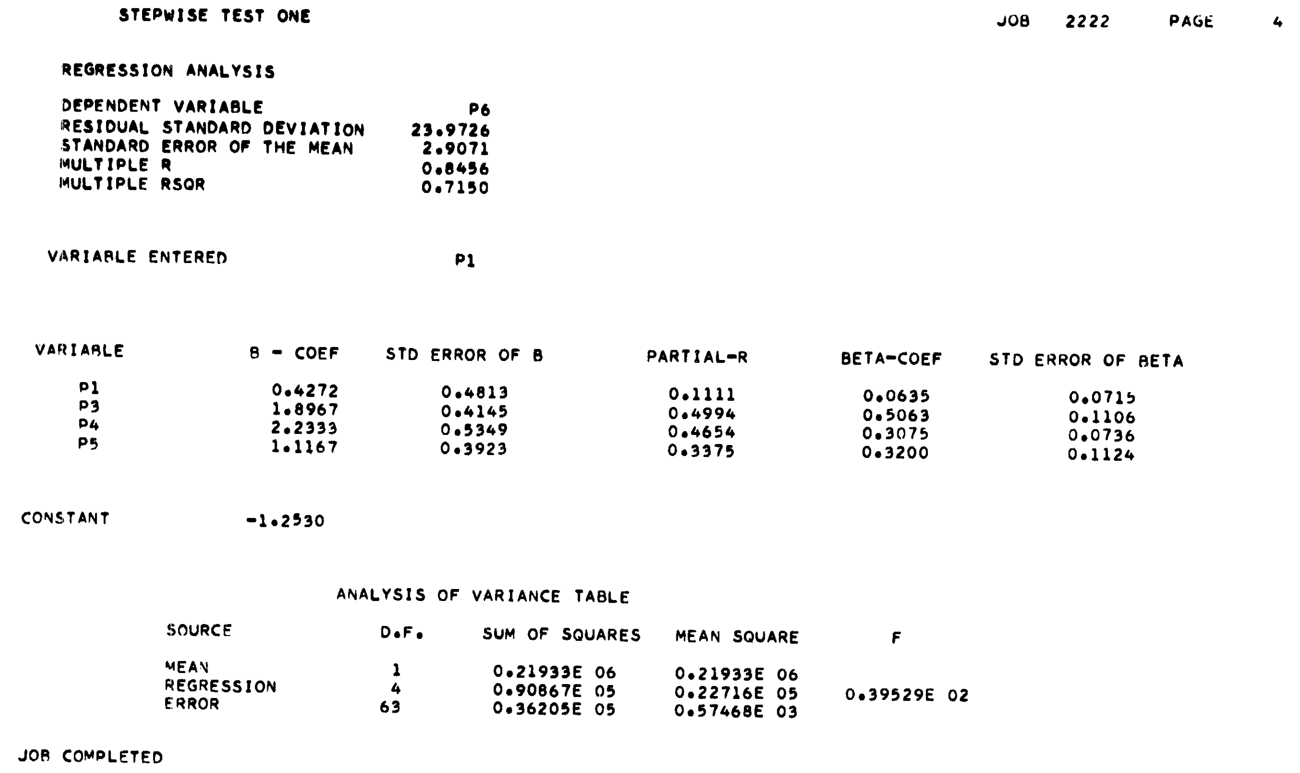
```
STEPWISE TEST ONE                                                  JOB    2222       PAGE       0

 NUMBER OF VARIABLES                    6
 INPUT TYPE                             3
 SEQUENCE CHECK                         0
 VARIABLES ON CARD 1                    0
 VARIABLES ON CARD 2                    0
 VARIABLES ON CARD 3                    0
 TRANSFORMATION SWITCH                  0
 OUTPUT RAW CROSS PRODUCTS              0
 OUTPUT RESIDUAL CROSS PRODUCTS         0
 PRINT PREDICTED VALUES                 0
 PRINT STEPS                            1
 POOLING OPTION                         0
 DEPENDENT VARIABLE                     6
 F-LEVEL TO REMOVE VARIABLES        0.500
 F-LEVEL TO ENTER VARIABLES         0.300
 TOLERANCE VALUE                  0.00010
 OUTPUT VARIANCE - COVARIANCE           0
 OUTPUT CORRELATION                     0
```

```
      STEPWISE TEST ONE                                                                   JOB    2222       PAGE        1

  REGRESSION ANALYSIS

  DFPFNDENT VARIABLE                   P6
  RFSIDUAL STANDARD DEVIATION     27.1238
  STANDARD ERROR OF THE MFAN       3.2892
  MULTIPLE R                       0.7860
  MULTIPLE RSQR                    0.6178


 VARIABLE ENTERED



 VARIABLE             B - COEF    STD ERROR OF B          PARTIAL-R         BETA-COEF    STD ERROR OF BETA

    P3                  2.9442        0.2850                0.7860           0.7860            0.0760


CONSTANT              26.0998


                                ANALYSIS OF VARIANCE TABLE

                SOURCE               D.F.      SUM OF SQUARES    MEAN SQUARE         F

                MEAN                   1        0.21933E 06      0.21933E 06
                REGRESSION             1        0.78516E 05      0.78516E 05     0.10672E 03
                ERROR                 66        0.48556E 05      0.73570E 03
```

```
      STEPWISE TEST ONE                                                                   JOB    2222       PAGE        2

  REGRESSION ANALYSIS

  DEPENDENT VARIABLE                   P6
  RESIDUAL STANDARD DEVIATION     25.0821
  STANDARD ERROR OF THE MEAN       3.0416
  MULTIPLE R                       0.8235
  MULTIPLE RSQR                    0.6781


 VARIABLE ENTERED                    P4



 VARIABLE             B - COEF    STD ERROR OF B          PARTIAL-R         BETA-COEF    STD ERROR OF BETA

    P3                  2.8275        0.2656                0.7971           0.7548            0.0709
    P4                  1.7976        0.5150                0.3972           0.2475            0.0709


CONSTANT              21.7445


                                ANALYSIS OF VARIANCE TABLE

                SOURCE               D.F.      SUM OF SQUARES    MEAN SQUARE         F

                MEAN                   1        0.21933E 06      0.21933E 06
                REGRESSION             2        0.86180E 05      0.43090E 05     0.68493E 02
                ERROR                 65        0.40892E 05      0.62911E 03
```



```
      STEPWISE TEST ONE                                                                  JOB    2222      PAGE      3

    REGRESSION ANALYSIS

    DEPENDENT VARIABLE                    P6
    RESIDUAL STANDARD DEVIATION      23.9328
    STANDARD ERROR OF THE MEAN        2.9022
    MULTIPLE R                        0.8435
    MULTIPLE RSQR                     0.7115


   VARIABLE ENTERED                     P5


  VARIABLE          B - COEF    STD ERROR OF B        PARTIAL-R        BETA-COEF    STD ERROR OF BETA

      P3              1.9583         0.4079            0.5145            0.5228            0.1089
      P4              2.3123         0.5266            0.4811            0.3184            0.0725
      P5              1.0355         0.3808            0.3217            0.2967            0.1091


 CONSTANT            2.9105


                         ANALYSIS OF VARIANCE TABLE

              SOURCE          D.F.      SUM OF SQUARES    MEAN SQUARE         F

              MEAN              1        0.21933E 06     0.21933E 06
              REGRESSION        3        0.90415E 05     0.30138E 05    0.52617E 02
              ERROR            64        0.36658E 05     0.57278E 03
```



```
      STEPWISE TEST ONE                                                                  JOB    2222      PAGE      4

    REGRESSION ANALYSIS

    DEPENDENT VARIABLE                    P6
    RESIDUAL STANDARD DEVIATION      23.9726
    STANDARD ERROR OF THE MEAN        2.9071
    MULTIPLE R                        0.8456
    MULTIPLE RSQR                     0.7150


   VARIABLE ENTERED                     P1


  VARIABLE          B - COEF    STD ERROR OF B        PARTIAL-R        BETA-COEF    STD ERROR OF BETA

      P1              0.4272         0.4813            0.1111            0.0635            0.0715
      P3              1.8967         0.4145            0.4994            0.5063            0.1106
      P4              2.2333         0.5349            0.4654            0.3075            0.0736
      P5              1.1167         0.3923            0.3375            0.3200            0.1124


 CONSTANT           -1.2530


                         ANALYSIS OF VARIANCE TABLE

              SOURCE          D.F.      SUM OF SQUARES    MEAN SQUARE         F

              MEAN              1        0.21933E 06     0.21933E 06
              REGRESSION        4        0.90867E 05     0.22716E 05    0.39529E 02
              ERROR            63        0.36205E 05     0.57468E 03

 JOB COMPLETED
```

2.2 PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

The aim of factor analysis is to explain observed relationships among numerous variables in terms of simpler relations. This simplification can take the form of producing a set of classificatory categories, or creating a smaller set of hypothetical variables.

The usual procedure is to collect measurements on n variables, over N persons or objects (N should be appreciably larger than n). To find out "what goes with what" among these n variables, the n variables can be intercorrelated as they vary over the N objects. This is done for all possible n(n-1)/2 pairings of the variables, producing a square symmetrical correlation matrix, R.

The process of factor analysis (or principal component analysis) is designed to resolve this correlation matrix into an n x k factor matrix, in which the number of factors, k, is usually considerably smaller than n, the number of variables. These factors may be considered as underlying influences which, in further measurement, can be substituted for the more numerous original variables, and which largely account for the correlations among the latter.

In analyzing the structure of a correlation matrix, two approaches can be taken. Formally, they resemble one another to a certain extent, but they have, in fact, rather different aims. One method is principal component analysis; the other is factor analysis. The former method is a relatively simple technique of "breaking down" a correlation matrix into a set of orthogonal (uncorrelated) components equal in number to the original variables. These correspond to the latent roots (eigenvalues) and accompanying latent vectors (eigenvectors) of the matrix. The method has the property that the roots are extracted in descending order of magnitude; this is important if only a few of the components are to be used for summarizing the data. These vectors are mutually orthogonal, and the components derived from them are uncorrelated. Although a few components may extract a large proportion of the total variance of the original variables, all components are required to reproduce the correlations between the variables exactly. Note that when the principal components method is employed, no hypothesis need be made about the original variables. They need not even be random variables. although, in practice, their values are usually regarded as a sample from some population.

Factor analysis, on the other hand, seeks to account for, or "explain", the matrix of correlations by a minimum, or at least a small number, of hypothetical variables or factors. Factor analysis asks the question, "Does a random variable F_1 exist such that the partial correlations between pairs of variables are zero after the effect of F_1 has been removed?" If the correlation matrix is still unexplained, the question is asked whether two random variables, F_1 and F_2, exist so that the partial correlations between pairs of variables are zero after the effects of both of these variables have been removed, etc. Thus, it may be said that principal components analysis is variance-oriented, whereas factor analysis is covariance-oriented.

As has been noted above, the number of factors needed to explain the correlations is fixed by the data itself, in the sense that when the factor extraction process leaves a residual correlation matrix of approximately zero, all of the covariance present has been accounted for.

However, this brings up the biggest problem in factor analysis. When a set of variables is intercorrelated, we have a set of n(n-1)/2 correlations. This leaves unanswered the question of what to put in the diagonal of the matrix, since we need a complete matrix for the process of factor analysis. Two solutions to this problem exist: (1) put ones in the diagonal — the method of principal components — on the grounds that, except for errors in measurement, a variable should correlate perfectly with itself; (2) insert values into the diagonal known as communalities (the term communality means the amount of variance of the variable accounted for by all the common factors together). This will obviously be less than the total variance, since some of the variance in any correlation matrix will be error variance, and some variance specific to that variable. This second solution is called factor analysis.

Using the method of principal components, it is possible to account for many variables by a few factors, since the first few principal components usually account for most of the variance. However, unlike the factor analysis model, the variance accounted for will include both specific factor and error variance. Factor analysis, in putting communality estimates, instead of ones, in the diagonal, attempts to partition the common factor variance from specific factor and error variance.

Unfortunately, use of the factor analysis model leaves the problem of deciding what values to use for communalities. The communality of a variable is the most that it has in common with other variables; thus, the squared multiple correlation of a variable with all of the other variables constitutes a lower bound on the communality. The true communality lies somewhere between the squared multiple correlation and one. To date, no method has been found of arriving at the "true" communality.

The problem is further complicated by the fact that the communality estimates and the number of factors extracted are mutually interdependent; the communality estimates chosen and the number of factors chosen determine when the residual correlation matrix drops to zero. Since, both of these values must be solved for simultaneously — and this is impossible — one has to start with one fixed and allow the other to be decided on by the program. Probably the best method to use is to fix the number of factors, and by iteration find communalities that exactly fit the off-diagonals to give that number of factors. In other words, decide on a number of factors, insert the squared multiple correlations as initial estimates of the communalities, and factor the correlation matrix, from which a new set of communalities is obtained; using these new values, refactor the matrix again. This process is continued until the change in communalities between successive factorings becomes trivial.

A final problem is deciding the number of factors to extract. In the absence of prior knowledge of the number of common factors in the correlation matrix, the safest course to adopt is Guttmann's lower bound theorem (1954), which demonstrates that eigenvalues with roots less than 1.0 are statistically insignificant, so that one could use this to set an upper bound on the number of factors to extract. Another possibility, which can be used in conjunction with the method of principal components, is to extract components until a prespecified amount of the total variance in the correlation matrix has been extracted. This can be done most successfully when there is some idea as to the amount of error variance in the correlation matrix (that is, the reliability of the measurements is known). Thus, a slightly smaller proportion of the variance can be extracted than is known to be common variance, since to take out more factors is to include error variance.

In addition to finding out the number of factors (or components) required to account adequately for an observed set of variables, we may also be interested in finding out, or defining, what these variables are. When a predetermined number of factors are extracted from a correlation matrix, we have a matrix of factor loadings, with k columns (for the k factors), and n rows for the n original variables. These factor loadings for variables v_1, v_2, $\ldots v_n$ on factors 1, 2, ...k are the correlations of the newly discovered factors with the original variables.

The concept of simple structure applies to the factor loading matrix. As has been pointed out above, the factor loading matrix consists of numbers such that each number corresponds to a given factor and a given variable. A particular element in the matrix indicates the extent to which that factor is represented in a given variable. However, the particular configuration of numbers obtained in an unrotated factor loading matrix is largely a function of the particular method used in extracting the latent roots and vectors of the correlation matrix, and may have no empirical meaning.

The concept of simple structure was developed as a number of criteria for the orthogonal rotation of the original factor loading matrix into such a position that the factors extracted are readily identifiable in terms of the original variables. Simple structure is a nonmathematical concept that sets up several criteria for rotation. The first of these is the existence of a positive manifold. This means that — all other things being equal — the factor loading matrix should have a minimum number of negative values. But for many factor loading matrices corresponding to particular correlation matrices, we may still have a very large number of factor loading matrices, which equally well account for the intercorrelations, and which all have mostly positive values. To further restrict the selection of the particular matrix, by which we shall define the primary variables, we require that each column of the matrix shall have a small number of high factor loadings, and a large number of near-zero loadings. We also require that each row of the factor loading matrix shall have at least one near-zero factor loading, and that at least one of the others shall be large-positive. In general, the fewer the number of large loadings and the larger the number of near-zero loadings in a row, the more simple the structure of that particular variable. The more variables that can be expressed in the simplest form in terms of a relatively large number of near-zero loadings, the more simple is the structure of the factor loading matrix.

However, the knowledge that there are a relatively large number of zeros in each row and column of the matrix of factor loadings is insufficient; the relative positions of the zeros and high loadings in the matrix must also be taken into account. For every pair of factors, the factor loadings should be arranged so that not many variables have high loadings on both. If variables have high loadings on the same factor, this implies that they tend to measure the same factor. Also, for any pair of factors, a number of variables should have high loadings on one factor and near-zero loadings on the other.

According to these criteria, if the loadings of one factor were to be plotted against another, we should find that some variables would cluster about zero, some would be high on one axis and low on the other, and vice versa for other variables. Thus, the procedure for finding the best simple structure matrix for any given set of variables would give an indication of which variables were the best measures for which factors. The factors would then be defined in terms of the variables that have relatively high loadings on them. In other words, simple structure is the application of Occam's Razor to the factor loading matrix, since it aims to explain the configuration of variables in such a way that each factor is represented by only a few variables (that is, it loads or correlates with the smallest possible number of variables).

Practically, simple structure is realized by several computational techniques. The best known is the Varimax rotation (Kaiser, 1958), which aims to maximize the fourth power of the factor loadings; this amounts essentially to maximizing the scatter among the loadings. Since a few highs means several lows, this leads to finding a position in which there are many low loadings. Thus, Varimax aims to prevent a variable being simultaneously highly loaded on two factors.

The Varimax rotation retains the property of orthogonality among the factors (that is, the factors remain uncorrelated). In other words, the clusters of variables that load highly on each of the factors are essentially uncorrelated with one another. In actual practice, this seems rather unlikely. It would seem more plausible that the clusters of variables (or factors), which are obtained from a factor analysis, are probably correlated (positively or negatively) with one another. It is for this reason that oblique simple structure rotations of the factor loading matrix are performed. The criteria for oblique simple structure are identical to those for orthogonal simple structure, with the exception that the orthogonality restriction is relaxed (that is, the factors need not be uncorrelated with one another). Basically, rotation to oblique simple structure is achieved by first performing a Varimax rotation, and then using the Varimax factor loading matrix to rotate obliquely. The effects of oblique rotation are usually to maximize the high loadings on each factor and minimize the near-zero loadings. The clusters of variables on each factor, as derived from the Varimax rotation, will not be altered; the effect simply is to give a ''cleaner'' solution.

When an oblique rotation is used, the situation is rather more complex than in the orthogonal case. There are, in fact, three matrices that must be understood in relating factors and variables:

1. The factor structure matrix, which gives the correlations between factors and variables

2. The factor pattern matrix, which gives the loadings of factors on variables

3. The factor estimate matrix, which gives the beta weights for estimation of factors from variables

The final feature is the estimation of factor scores, for both the orthogonal and the oblique case. When a given number of factors have been extracted from a large number of variables, and they have been identified by rotation of the factor loading matrix, it may be desirable to calculate, for each observation, a set of factor scores, in place of the scores on the original variables. These scores are a weighted summation of the original scores on each of the variables for a given observation. These scores can then give a profile score for each observation over the factors extracted.

2.2.1 Mathematics of Principal Components and Factor Analysis

Principal Components Analysis

Let us consider n points in a space of p dimensions, when the x's are expressed in standard score form (that is, mean = 0; variance = 1). The line with current coordinates X is

$$\frac{X_1-m_1}{g_1} = \frac{X_2-m_2}{g_2} = \dots \frac{X_p-m_p}{g_p} \qquad (1)$$

where g's are direction cosines, and are subject to the condition

$$\sum_{i=1}^{p} g^2_{\ i} = 1 \qquad (2)$$

The sum of squares of the distances from the n points on to this line is nS, where

$$nS = \sum_{j=1}^{n} (\sum_{i=1}^{p} (x_{ij}-m_i)^2 - (\sum_{i=1}^{p} g_i\ (x_{ij}-m_i))^2 \qquad (3)$$

If this is a stationary value, the partial derivatives with respect to the m's vanish, and

$$-\sum_{j} (x_{ij}-m_i) + \sum_{j} g_i \sum_{i} g_i (x_{ij}-m_i) = 0, \ i = 1, 2, .. p \qquad (4)$$

and since $\sum_{j} x_{ij} = 0$, $m_i/g_i = k$ (a constant).

Thus, the origin lies on the line (1), and we may take all the m's to be zero, thus

$$\sum_{j=1}^{n} x^2_{ij} = n, \qquad \text{and we have}$$

$$nS = \sum_{j=1}^{n} (\sum_{i=1}^{p} x^2_{ij} - (\sum_{i=1}^{p} g_i x_{ij})^2)$$

$$= np - \sum_{j=1}^{n} (\sum_{i=1}^{p} g_i x_{ij})^2 \qquad (5)$$

Then we can find stationary values of S for variations in g, subject to equation (2) above. If, we consider λ , an undetermined multiplier, this gives us

$$- 1/n\sum_{j=1}^{n} x_{kj}(\sum_{i} g_i x_{kj}) + \lambda g_k = 0, \quad k = 1, 2, \ldots p \tag{6}$$

which gives us the set of p equations

$$\begin{aligned} &g_1(1-\lambda) + g_2 r_{12} + \ldots g_p r_{1p} = 0 \\ &\cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ &\cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ &g_i r_{p1} + g_2 r_{p2} + \ldots + g_p(1-\lambda) = 0 \end{aligned} \tag{7}$$

Eliminating the g's, we get the characteristic equation of the correlation matrix

$$\left| r - \lambda I \right| = 0 \tag{8}$$

For a known r, this gives p roots in λ. To each root corresponds a set of g's for which S has a stationary value. Also, using equation (6), we find from equation (5)

$$S = p - \lambda \tag{9}$$

It follows from equation (9) that the root which gives the minimum S is the one with the largest λ . If we choose the largest root of equation (8), we have the line required. The sum of squares of distances of the points from it is a minimum, and the variate measured along it has the maximum variance. The variate is given by

$$V_1 = \sum_{j=1}^{p} g_{1j} x_j \tag{10}$$

This indicates that the set of g's relates to λ_1. If we multiply, equation (6) by x_k, and sum over k, we can see that the variance of V_1 is λ_1.

If we now look for the direction (perpendicular to the first line), for which the sum of squares of perpendiculars is at a minimum, we find the line corresponding to λ_2, the second largest root, etc.

Thus, we have transformed to a new set of variates, V, that are uncorrelated, and have variances λ_1, λ_2, ... λ_p in decreasing order.

Note that $p = \Sigma \lambda$, since the sum of the roots is the sum of p units in the main diagonal. Also, if the original variables are normally distributed, we can regard the V's as splitting off independent components of variance λ_1, λ_2, ... λ_p, from the total variable, p.

Thus, we can select from this set of p variates, V, the first n components and consider these as our factors. In general, the two main criteria for deciding on the number of components to retain are (a) when the value of the λ's falls below 1.0; or (b) when the first n components account for a fixed percentage of the total variance.

Factor Analysis Model

Basically, the mathematics and computational steps in factor analysis are similar to those required by the method of principal components. The difference lies in the fact that in component analysis we begin with a set of observations and look for components, in the hope that we shall be able to effect a reduction in the dimensions of variation, and that we can give some physical meaning to the components thus extracted. Using the factor analytic model, we begin with a theoretical model, and try to find out whether it agrees with the data, and if it does, to estimate its parameters.

Let us begin, as in the method of principal components, with a matrix of observations x_{ij}, and consider whether they can arise from a situation with the following structure:

$$x_i = \sum_{k=1}^{p} a_{ik} f_k + b_i s_i + c_i e_i, \qquad \text{where } i = 1 \ldots p \tag{1}$$

In this equation (1), the f_k are factors that can appear in more than one x, s_i is a factor specific to the variable x_i, and e_i is an error term.

At this level, the model is undetermined, and, in fact, by using the method of principal components, we can always express the x's in terms of f's without invoking specific or error terms at all. This is the basic difference between principal components and factor analysis. In the former, we consider all the variance, common variance, and extract its orthogonal components. In factor analysis, on the other hand, we take into account that some of the variance is going to be due to error, and some to variance that is quite specific to a certain variable. In this sense, the factor analytic model is more realistic, and the principal components method, in spite of its mathematical simplicity is misleading.

Therefore, let us assume that we have n observations on p variables x_i. Using the method of principal components, we have

$$S = \sum_i \text{var } x_i - \sum_i g_{ij} g_{ik} \text{cov } (x_j, x_k) \tag{2}$$

Thus, the principal components equations are

$$\sum_j {}_{ij} \text{cov}(x_j, x_k) = \lambda \ {}_{ig\ ik} \tag{3}$$

However, in the factor analysis model, if

$$x_{ij} = \sum_k a_{ik} f_{kj}$$

$$\mathrm{cov}(x_i, x_j) = \sum_{k,m} a_{ik} a_m 1/n \sum_j f_{kj} f_{mj} \tag{4}$$

and if we substitute its expected value for $\Sigma f_{kj} f_{mj}$

$$\begin{aligned} d_{km} &= 1, \qquad && \text{if } k = m \\ &= 0, && \text{if } k \neq m \end{aligned}$$

we have

$$\mathrm{cov}(x_i, x) = \sum_k a_{ik} a_k \tag{5}$$

If the model is

$$x_{ij} = \Sigma a_{ik} f_{kj} + e_{ij} \tag{6}$$

we have

$$\mathrm{cov}(x_i, x_j) = \Sigma a_{ik} a_{jk} + \mathrm{var}\ e_i d_i \tag{7}$$

Now we are required to estimate the coefficient a_{ij}. We can operate on the estimated matrix

$$\left| \Sigma a_{ik} a_k + d_i \mathrm{\ var\ } e_i \right|$$

This is the same as the former matrix, except for the principal diagonals, where each term is increased by var e_i. Thus we would like to have in the main diagonal not Σa_{ik}^2 + var e_i, but only Σa_{ik}^2. In other words, if we are not to bias the estimates of the a's, we must remove var e from the diagonal terms. This is equivalent to substituting communalities for unity in the diagonals of the standardized matrix.

Possibly the best way of estimating the communalities is to start with the squared multiple correlations in the diagonal (since this is a lower bound on the communalities). We then perform an analysis of the data and arrive at certain factors (deciding on the number of factors, by taking those with latent roots >1.0, those factors accounting for a certain percentage of the variance, or some other method). We can then use the coefficients occurring in those factors to estimate new communalities, iterate, and proceed until the communalities converge. Basically, what this process amounts to is that we assume m factors and assume that they account for as much as possible of the variance; this determines the communalities and, consequently, the "error" variances. But this does not mean that we have estimated the actual error variances that occur in practice. We have estimated only what they would be if the number of factors is what we think it is, and the error variances are minimal.

In other words, this would suggest that in practice care should be exercised in computing a factor analysis. If one has no idea of the number of factors to extract, possibly the best solution is to compute a principal components solution of various numbers of factors, rotate, and decide which set of factors gives the best empirical meaning. Then, using this number of factors, estimate the communalities, and run a proper factor analysis.

Rotation of the Factor Loadings

In using the Varimax criterion for rotation to orthogonal simple structure, we have already defined verbally what we mean by simple structure. Mathematically, the simplicity of the factorial composition of the jth variable can be defined as the variance of the squared loadings for the test

$$q^*_j = (r\sum_s (a^2_{js})^2 - (\sum_s a^2_{js})^2)/r^2 \qquad (1)$$

where j = 1, 2, ... n variables, and s = 1, 2, ... r factors, and a_{js} is the factor loading of the jth variable on the sth factor.

To obtain the total criterion for the entire factor matrix, we can sum over the variables thus:

$$q^* = \sum_j \; (r\sum_s (a^2_{js})^2 - (\sum_s a^2_{js})^2)/r^2 \qquad (2)$$

This criterion can be modified if we define the simplicity of a factor as the variance of its squared loadings

$$v^*_s = (n\sum_j \; (a^2_{js})^2 - (\sum_j a^2_{js})^2)/n^2 \qquad (3)$$

and for the criterion for all the factors, define the maximum simplicity of a factor matrix as the maximization of

$$v^* = \sum_s v^*_s = \sum_s ((n\sum_j (a^2_{js})^2)/n^2$$

which is the variance of squared loadings by columns rather than by rows.

This is the row Varimax rotation. However, this will exhibit a systematic bias because of the divergent weights that implicitly are attached to the variables by their communalities. Therefore, the normalized Varimax rotation weights each variable by its communality, thus:

$$v = \sum_{s} ((n \sum_{j} (a_{js}^{2}/h_{j}^{2})^{2} - (\sum_{j} (a_{js}^{2}/h_{j}^{2}))^{2})/n^{2}) \quad (4)$$

where h_j^2 is the communality of the $j^{\underline{th}}$ variable. In this case, the variance of the squared correlations of the common parts of the variables with a factor are now being maximized.

The oblique rotational scheme is developed from the above normalized Varimax solution. The Promax rotation simply takes the Varimax-rotated factor loading matrix and generates a pattern matrix from it by powering all the elements in the original matrix.

We can define a matrix $P = (p_{ij})$ such that

$$p_{ij} = \left| a_{ij}^{k+1} \right| / \, a_{ij} \quad (5)$$

with $k > 1$. Each element of this matrix is, except for the sign that remains unchanged, the $k^{\underline{th}}$ power of the corresponding element in the row-column normalized orthogonal matrix. We then find the least-squares fit of the orthogonal matrix of factor loadings to the pattern matrix generated by equation (5).

$$L = (G'G)^{-1}G'P \quad (6)$$

where L = the unnormalized transformation matrix of the reference vector structure, G = the orthogonal-rotated matrix, and P = the matrix derived by equation (5) above. Finally, the columns of L are normalized so that their sums of squares are equal to unity.

Computation of Factor Scores (Direct Estimation)

Factor scores can be computed for each observation on the factors extracted from the original correlation matrix. (See Harman, 1960, Chapter 16, in list of references at end of this chapter.) These are computed using the following equations:

Let F = matrix of factor loadings
V = the orthonormal matrix of eigenvectors
A = diagonal matrix of latent roots (eigenvalues)

Then

$$F = VA^{1/2} \quad (1)$$

F is an orthogonal matrix (but not orthonormal). Since V is orthonormal

$$V' = V^{-1} \tag{2}$$

From the principal components model we know

$$Z = BS + US + eS \tag{3}$$

where Z = vector of scores for each observation
B = matrix of common factor loadings
U = diagonal matrix of uniqueness $(1-h_i^2)$
e = diagonal matrix of errors

Therefore

$$S = F^{-1}Z \tag{4}$$

since

$$\begin{aligned} F^{-1} &= V^{-1}A^{-1/2} \\ &= VA^{-1/2} \end{aligned} \tag{5}$$

so

$$F^{-1} = A^{-1}F' \tag{6}$$

$$S \quad = A^{-1}F'Z \tag{7}$$

Therefore, the factor score for an observation on factor 1 would be

$$s_1 = f_{11}Z_1/\lambda_1 + f_{12}Z_2/\lambda_1 + \ldots + f_{1p}Z_p/\lambda_1 \tag{8}$$

TERMS USED IN FACTOR ANALYSIS

Communality. Sum of squares of factor loadings for any given variable (that is, the total variance due to factors which this variable shares with other variables in the matrix).

Covariance. Mean product of deviations of variable x and variable y from their means $(1/n)\ \Sigma\ (x-\bar{x})\ (y-\bar{y})$.

Diagonal matrix. A square matrix having zeros in all positions except those on the diagonal from upper left to lower right.

Direction cosine. One of a set of cosines of angles, defined for a point, each angle being measured between one of the reference axes and the vector connecting the point with the origin.

Factor loading. Correlation of any particular variable with the factor being extracted.

Factor matrix. Matrix whose entries are the factor loadings obtained from a factor analysis; it generally is arranged so that it has as many columns as factors extracted, and as many rows as variables.

Hyperplane. Space of (N-1) dimensions, defined by a reference vector perpendicular to it. (For example, in two dimensions, either coordinate axis is the hyperplane of the other; in three dimensions, the plane defined by any two coordinate axes is the hyperplane of the third.)

Normalize. To divide each of a set of numbers by the square root of the sum of squares of all numbers in the set, so that the sum of squares of the new set is 1.00.

2.2.2 The Program

Given a set of observations numbering 499 or fewer containing measures on $n \leq 30$ variables, x_1, x_2, $\ldots x_n$, a square symmetric n x n correlation matrix, R, and vectors of means and standard deviations are computed. From this or a given correlation matrix, a factor matrix is extracted containg n or fewer vectors of factor coefficients. This factor matrix may be rotated to approximate simple structure by an analytical criterion in either an orthogonal or an oblique reference frame. Given the original data, X, together with its means and standard deviations, this rotated factor matrix, G, may be used to compute factor measurements (factor scores) by a regression model.

Communalities, h_i^2, are treated as the diagonal elements of the correlation matrix. These elements are computed equal to 1.0, and should be retained as such for the computation of the principal components factor matrix. They may be specified by some predetermined values, however, where a specially constructed correlation matrix is given rather than carried over from previous computation. For example, the user may desire to place specific communality values on the diagonal of this matrix (see section 2.2.6). Options are also available to consider each $i^{\underline{th}}$ communality as the maximum absolute off-diagonal element in the $i^{\underline{th}}$ vector of the correlation matrix or as the squared multiple correlation for the $i^{\underline{th}}$ variable.

All n of the latent roots of the correlation matrix with diagonal elements chosen as indicated above are computed by a Householder (HOW, 1962) tridiagonalization, followed by the use of the QR algorithm. The k latent vectors are computed by Wilkinson's (HOW, 1962) method. These latent roots (eigenvalues) are solutions to the matrix equation

$$(R-\lambda)\ A^* = 0$$

where R is the correlation matrix, λ is a diagonal matrix of the latent roots, and A^* is a matrix of the latent vectors (eigenvectors).

The total variance accounted for by the principal components of the correlation matrix is evaluated by its trace (the sum of its diagonal elements). The percentage of this total variance accounted for successively by each latent root is computed and presented cumulatively for the first through the $n^{\underline{th}}$ root. This percentage is presented whether or not a principal components analysis ($r_{ii} = 1$) is being performed. If $r_{ii} \neq 1$, the output should be ignored. The latent vectors, a^*_j, are normalized, and a matrix, F, of factor loadings is computed by scaling each latent vector by the square root of its associated latent root, that is,

$$f_{ij} = a_{ij}\sqrt{\lambda_j}\ ,$$

where the f_{ij} are factor loadings, the a_{ij} are elements of the latent vectors, and the λ_j are latent roots.

Several methods are available to the user for estimating the rank, m, of the factor space for the purpose of retaining only m factors for output or subsequent rotation. The value of m may be specified on a control card and arbitrarily accepted as the maximum number of factors. The user may also request that only those factors whose latent roots are equal to or greater than 1.0 be retained. An option is also available to retain only those factors which cumulatively account for an amount of variance equal to or less than a given percentage of the total variance. This option could be used in a principal components analysis.

The matrix of factor loadings may be rotated to approximate simple structure in an orthogonal reference frame by the Normal Varimax method (Kaiser, 1958) for the case of uncorrelated factors; k factors are rotated where k is equal to or less than m, the rank of the factor space as determined above. The Normal Varimax method develops a transformation matrix, T, over a cycle of rotations of each of the $\frac{k}{2}(k-1)$ pairs of orthogonal axes of the factor space taken in turn. The angle of each rotation is chosen such that a function, U, of the factor matrix is maximized. Complete cycles of rotations are performed until U is not significantly increased by an additional cycle. U is computed by an evaluation of the following expression:

$$U = k\sum_{i=1}^{n}\sum_{j=1}^{k}\left[\frac{g_{ij}}{h_i}\right]^4 - \sum_{j=1}^{k}\left[\sum_{i=1}^{n}\frac{g_{ij}^2}{h_j^2}\right]^2$$

where k is the number of factors, n the number of variables, and g_{ij} an element of the factor matrix under rotation for the $i^{\underline{th}}$ variable on the $j^{\underline{th}}$ factor; h_i^2 represents the communality of the $i^{\underline{th}}$ variable computed using only the k factors under rotation. The final rotated factor matrix, G, is derived by the matrix multiplication, G = FT, where T is the complete Varimax transformation matrix, and F is the unrotated factor matrix.

This matrix of k factors may also be rotated to approximate simple structure in an oblique reference frame by the Promax method (Hendrickson and White). A pattern matrix, P^* describing a factor matrix rotated to approximate simple structure in oblique axes, may be accurately estimated by a matrix whose elements are functions of the elements of the orthogonal matrix rotated by the Varimax method. This matrix can be derived by the following operation:

$$P^*_{ij} = \frac{\left| g_{ij}^{a+1} \right|}{g_{ij}}$$

where P^*_{ij} is an element of the pattern matrix, P^* and g_{ij} is an element of the orthogonally rotated factor matrix, G. The value of a is four (4). According to Hendrickson and White (1964), four is the optimal value in the majority of cases. The user can, however, easily change this number in the program PROMX. In general, as a increases, the dependence of the rotated factors or their obliquity increases. A transformation matrix is computed that rotates the orthogonal factor matrix into an oblique reference vector structure matrix, V, which is a least-squares fit to the pattern matrix, P^* described above. This transformation matrix, L, is derived in unnormalized form from the following matrix equation:

$$L = (G'G)^{-1}G'P^*$$

After the transformation matrix, L, has been column-normalized, the reference vector structure matrix, V, is obtained from $V = GL$. The correlation matrix of the reference vectors, ψ, is computed by $\psi = L'L$, and the reference vector pattern matrix, W, is then developed by $W = V\psi^{-1}$. To derive the primary factor structure from the reference vector solution described above, the diagonal matrix, D, of the correlations among the reference vectors and the primary factors is computed by taking

$$d_{ii} = \frac{1}{\sqrt{c_{ii}}}$$

where the d_{ii} are diagonal elements of D, and c_{ii} are diagonal elements of the matrix, ψ^{-1}. The primary factor structure matrix, S, is then determined by $S = WD$, and the primary factor pattern matrix is derived by $P = VD^{-1}$. The matrix of correlations between the primary factors, Φ, is computed by $\Phi = D\psi^{-1}D$.

Factor measurements (factor scores) are computed by the "short" regression method (Harman, 1960). The diagonal matrix of uniquenesses, U, is obtained by taking $u_{ii} = 1-h^2_i$, where the h^2_i are the communalities computed from G, the orthogonal factor matrix, or P, the oblique factor matrix, depending on which scores are requested. For the case of uncorrelated or orthogonal factors, the matrix, Q, is developed by $Q = I + G'U^{-1}G$, where G is the orthogonal factor matrix (Varimax solution). Factor scores are formed by the operation $\bar{f} = \beta' Z$, where $\bar{f}$ is a factor score matrix, β is a matrix of factor score regression coefficients, and Z the vector of standardized data. β is obtained by the operation $\beta' = Q^{-1}G'U^{-1}$. The elements of Z are computed by

$z_{kj} = \frac{x_{kj} - \bar{x}_j}{s_j}$, where x_{kj} is an observation in a data matrix for the k$^{\underline{th}}$ sample case on the j$^{\underline{th}}$ variable, and $\bar{x}_j$ is the mean and s_j the standard deviation of the j$^{\underline{th}}$ variable.

For the case of oblique or correlated factors, the procedure is much the same, except for the definition $Q = \Phi^{-1} + P'U^{-1}P$, where P is the primary factor pattern matrix and Φ the matrix of intercorrelations of the primary factors. Also, here, $\beta' = Q^{-1}P'U^{-1}$.

2.2.3 Summary of Output Statistics

1. High and low value of each variable
2. Means of each variable
3. Standard deviation of each variable
4. Sample variance for each variable
5. Matrix of raw cross products
6. Matrix of residual cross products
7. Variance-covariance matrix
8. Matrix of correlation coefficients
9. Matrix of characteristic vectors
10. Characteristic values
11. Trace
12. Cumulative percentage of trace of each characteristic value
13. Unrotated factor matrix
14. Orthogonal transformation matrix
15. Orthogonal factor matrix
16. Transformation matrix to oblique reference vector structure
17. Oblique reference vector structure matrix
18. Correlations among oblique reference vectors
19. Oblique reference vector pattern matrix
20. Oblique primary factor structure matrix
21. Correlations among oblique primary factors
22. Oblique primary factor pattern matrix
23. Factor score regression coefficients
24. Factor scores
25. Communalities

2.2.4 Job Execution

To perform a factor or principal components analysis, the user must supply three sets of cards to the program:

1. Monitor control cards
2. Program control cards
3. Data cards

Descriptions of the form and content of each card set follow.

MONITOR CONTROL CARDS

The monitor control cards are necessary to initiate program loading from the disk and to establish the necessary communication with the monitor. A general description of the cards may be found in IBM1130 Disk Monitor Reference Manual (C26-3750).

A factor analysis requires the following monitor cards:

```
CC: 1  4   8         16-17
       ↓   ↓           ↙
    // XEQ FCTR      04

    *LOCALFCTR, FMTRD, DATRD, PRNTB, MXRAD, TRAN
    *LOCALFCTR1, TRIDI, QR, INVRS
    *LOCALFCTR2, VECTR, PRNT
    *LOCALFCTR3, VARMX, PROMX, SCORE, RFOUT
    *LOCALCOREL, PRNT
```

The monitor control cards do not change from job to job within one analysis, but must be included with every job processed. The first program operated on by this system should be preceded by a cold start card.

PROGRAM CONTROL CARDS

The program control cards communicate the data-specific parameters and output options to the program. There are five possible card types necessary for execution:

1. Input/output units card*
2. Job-title card*
3. Option card (described below)
4. Variable name card (described below)
5. Variable format card*

Four of the control cards are required in every job. The variable format card is necessary only if source data is to be processed.

*See "General Operating Instructions", section 1.2.

OPTION CARD

Number of Variables (cc 1-2)

This field must be punched with a nonzero integer, n, which is less than or equal to 30. The value contained in n is the total number of variables tobe processed.

Input Type and Source (cc 3-4)

This field allows the user to specify the input device (1442 card reader or disk) and, indirectly, the type of input analysis to be undertaken in the input program. The three possible values that may be punched in this field are described below:

Value	Meaning
1	Raw data will be read from the 1442 card reader and transferred to the disk. It will be retained there for use by this or other programs until destroyed by input from one of the four programs in this system. Raw sums and raw sums of cross products will be accumulated. Data will be read until a card with a negative number in the identification field is encountered (section 2. 2. 5).
2	Raw data will be read from the disk. Raw sums and raw sums of cross products will be accumulated. Data will be read until a negative integer in the identification field is encountered.
3	A previously computed matrix, or matrices, will be read from the 1442 card reader. Matrix cards will be read until a negative job number field is encountered (see "Pooling", section 2.2.6).

Sequence Checking (cc 5-6)

This field is used to indicate that raw data input from the card reader (cc 3-4 contains a 1) is to be sequence-checked. A value of zero or a blank field implies that no sequence check will be made. A value of one (1) implies that the cards will be sequence-checked. The sequence-checking process consists of an equal comparison check of the case identification field for all cards in a case and an ascending sequence check of the card number field. If an error in either of these conditions is encountered, the program prints a message, and the job is terminated.

Number of Variables on Card 1 (cc 7-8)

When a data vector contains more variables than will fit on one card, the user must indicate to the program the number of variables punched on each card. This field must be punched with the number of variables on the first card. If there is only one card per case, this field must be blank or zero.

Number of Variables on Card 2 (cc 9-10)

Same as cc 7-8, except that this field indicates the number of variables on the second card of the data.

Number of Variables on Card 3 (cc 11-12)

Same as cc 7-8, except that this field indicates the number of variables on the third card of the data.

Transformation Switch (cc 13-14)

If the value in this field is nonzero, a user-written transformation subroutine is called after each data record is read and before any computation takes place.

If the value in this field is zero or blank, the transformation subroutine is not called.

The use of transformations is discussed in section 2.5.1.

Output Raw Sums of Cross Products (cc 15-16)

This field is used to indicate whether the raw sums and sums of raw cross products matrix are to be printed, punched, printed and punched, or not presented.

The four (4) possible values of this field are described below. The computation to generate the matrix is performed even if the "no output" option is chosen.

Value	Meaning
0 or blank	No output.
1	Matrix will be printed.
2	Matrix will be printed and punched.
3	Matrix will be punched.

Punched output of the raw sums of cross products matrix includes the number of observations and the vector of raw sums and sums of squares. This entire output must be entered on the pooling option (section 2.2.6).

Output Residual Cross Products (cc 17-18)

This field is used to indicate whether the residual cross products matrix — defined as:

$$u_{ij} = c_{ij} - \frac{s_i s_j}{n} \qquad i, j = 1, 2 \ldots n$$

where c_{ij} are the elements of sums of raw cross products matrix, and s_i, s_j are the raw sums of the $i^{\underline{th}}$ and $j^{\underline{th}}$ variables, respectively, and n is the number of cases — is to be printed, punched, printed and punched, or not presented.

The four (4) possible values are described above under "Output Raw Sums of Cross Products".

The matrix is computed even if the "no output" option is chosen.

Output Variance-Covariance Matrix (cc 19-20)

This field is used to indicate whether the variance-covariance matrix — defined as:

$$c_{ij} = \frac{u_{ij}}{n-1} \qquad i, j = 1, 2 \ldots n$$

where u_{ij} is an element of the residual cross products matrix, and n is the number of cases — is to be printed, printed and punched, punched, or not presented. The four (4) possible values that may occur in this field are as is given for the above matrices.

There are no additional vectors or matrices punched with the punched output. The matrix is computed even if the "no output" option is chosen.

Output Correlation Matrix (cc 21-22)

This field is used to indicate whether the correlation matrix — defined by:

$$r_{ij} = \frac{c_{ij}}{s_i s_j} \qquad i, j = 1, 2 \ldots n$$

where c_{ij} is an element of the variance-covariance matrix, and s_i, s_j are the standard deviations of the $i^{\underline{th}}$ and $j^{\underline{th}}$ variables, respectively — is to be printed, punched, printed and punched, or not presented. The four (4) possible values contained in this field are as is given for other matrices, above.

The punched output of the correlation matrix includes the number of cases and cards containing the vectors of means and standard deviations.

The matrix is generated even if the "no output" option is chosen.

Factor Scores (cc 23-24)

This field is used to indicate whether factor scores are to be computed. If a value of zero (0) is punched or the field is left blank, the factor score computation is suppressed. When this field contains a one (1), factor scores and factor score regression coefficients are computed. A two (2) in this field causes scores to be punched. Upon entry to the program SCORE, which computes the scores, the program assumes that the necessary matrices and data have been set up for the analysis. Hence, either an orthogonal or an oblique rotation must be performed before entry to the SCORE routine. The number of scores to compute is equivalent to the number of rotated factors. Hence, it is not possible to compute factor scores from the unrotated principal axis factor matrix. In addition to the factor matrix and an auxiliary matrix computed by the rotation output program, the scores program requires the data file to be on the disk and the means and standard deviations vectors to be located in common storage. If the entire factor analysis is being done from the raw data, these operations are performed automatically by the program.

Factor Score Punched Output Format:

Card I

Columns	Elements
1-4	Factor score number (I)
5-6	25 (Identifier)
7-20	Score for variable 1; ± 0 .XXXXXXXE±XX
21-34	Score for variable 2; "
.	.
.	.
.	.
63-76	Score for variable 5; "

If more than five factors have been rotated, card I + 1 has a format as follows:

Columns	Elements
1-6	Blank
7-20	Score for variable 6
.	.
.	.
.	.

The factor scores are computed from the Varimax solution if an oblique rotation has not been requested; otherwise, they are computed from the oblique solution.

Number of Factors to Compute (cc 25-26)

The number punched in this field is used as a switch setting in the program to determine the number of factors to compute from the characteristic roots and vectors. There are four (4) possible values that may be punched, and they are described below:

Value	Meaning
0 or blank	No factors will be computed.
1	Only those factors will be computed whose characteristic vectors have associated characteristic roots greater than or equal to one (1).
2	Compute a fixed number of factors (m). The value of m will appear in cc 27-28; m must not be greater than ten.
3	Compute factors whose variance accounts jointly for no more than P percent of the total variance. The value of P appears in cc 27-28. The variance of a factor is the characteristic root associated with a particular characteristic vector. The

Value	Meaning
3 (cont)	total variance is defined as the trace of the matrix. The percentage is computed by adding characteristic roots and forming the ratio of this sum to the trace of the communality-adjusted correlation matrix.

Constant for Number of Factors (cc 27-28)

This field is used in conjunction with cc 25-26. A two (2) in the previous field implies that this field will contain an integer, m, which is equal to the number of factors to compute. If cc 25-26 contains a three (3), this field should contain an integer, P, which is the percentage of factor variance.

Communality Estimation Options (cc 29-30)

This field is used to offer the user a choice of three methods of estimating the communality or common variance of the reduced factor space. Before the characteristic roots and vectors are computed, the program places the communality estimate on the principal diagonal of the matrix to be factored. Three possible values may be punched in this field, and they are described below. Iteration on communalities is not performed automatically, and is discussed in section 2.2.6.

Value	Meaning
0	No change to the matrix.
1	The absolute value of the largest off-diagonal element in a row will be used as the communality estimate for that variable.
2	The square of the multiple correlation coefficient between variable i and all other variables in the matrix will be used as the communality estimate for the ith variable. This is done for all i.

Rotation Switch (cc 31-32)

This field is used to indicate the type of rotation to simple structure that will be used on the principal axis factor matrix. The user has the choice of choosing an orthogonal rotation (normal Varimax) and/or an oblique rotation (Promax). The process by which an oblique rotation is computed requires that an orthogonal rotation be performed first. Hence, in addition to the oblique rotation matrices, the user has the option of obtaining the output of the orthogonal rotation. Three possible values may be punched in this field, and they are described below:

Value	Meaning
0 or blank	No rotation will be performed.
1	Orthogonal rotation.
2	Oblique rotation (includes an orthogonal rotation).

Number of Factors to Rotate (cc 33-34)

This field is used in conjunction with the rotation switch described in the previous field. The value punched in this field determines the number of factors to rotate. Two possible conditions can arise. If the user does not know the number of factors to rotate, it is suggested that the field be left blank (or zero). The number of factors to rotate is then chosen on the basis of one of the options in cc 25-26. However, if a value of k appears, k factors are rotated if this number is less than or equal to the number of factors computed. In any case, k must be less than or equal to ten.

Pooling Option (cc 35-36)

When using the matrix input/output option (03 in cc 3-4) and when pooling sums of squares and cross products (section 2.1.4), if the user desires that matrices be subtracted rather than added, this field should be nonzero.

Factor Matrix Output Option (cc 37-62)

In a complete factor analysis, there are 13 additional matrices (section 2.5.3) that the user has the option to output, if desired. The 13 remaining fields on the card are for this purpose.

Each of the two-column fields may take four possible values, described below:

Value	Meaning
0 or blank	No output.
1	Print matrix.
2	Print and punch matrix.
3	Punch matrix.

The following gives the name and field column numbers for each matrix:

Column	Matrix
37-38	Characteristic vectors, A*
39-40	Unrotated factors, F
41-42	Orthogonal transformations, T
43-44	Orthogonal factors, G
45-46	Transformation to oblique reference vector structure, L
47-48	Oblique reference vector structure, V
49-50	Correlations among oblique reference vectors, ψ
51-52	Oblique reference vector pattern, W
53-54	Correlations between reference vectors and primary factors, D
55-56	Oblique primary factor structure, S
57-58	Correlations among oblique primary factors, Φ
59-60	Oblique primary factor pattern, P
61-62	Factor score regression coefficients, β

Factor Analysis Option Card Summary

Column	Meaning
1-2	Number of variables
3-4	Input type and source 1 - Raw data input from card reader 2 - Raw data input from disk 3 - Matrix input from card reader
5-6	Check sequence of raw data input 0 - No 1 - Yes
7-8	Number of variables on card 1
9-10	Number of variables on card 2
11-12	Number of variables on card 3
13-14	Transformation switch 0 - No transformation 1 - Transformation
15-16	*Output raw cross products matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch
17-18	*Output adjusted cross products matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch
19-20	*Output variance-covariance matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch
21-22	*Output correlation matrix 0 - No 1 - Print 2 - Print and punch 3 - Punch

* Not available when correlation matrix is used as input.

Column	Meaning
23-24	Factor score options 0 - Do not compute factor scores. 1 - Compute and print factor scores. 2 - Compute, print, and punch factor scores.
25-26	Number of factors options 0 - Do not compute factors. 1 - Compute factors for latent roots ≥1.0 only. 2 - Compute m factors (where m is given in cc 27-28). 3 - Compute factors accounting jointly for no more than p percent of the total variance (where p is given in cc 27-28).
27-28	Constant for number of factors option, if appropriate
29-30	Communality options 0 - Use diagonal values of correlation matrix (normally unity unless otherwise specified in a given matrix). 1 - Use maximum absolute off-diagonal element in each vector of the correlation matrix. 2 - Use the squared multiple correlation coefficient for each variable.
31-32	Rotation options 0 - Do not perform any rotations. 1 - Perform an orthogonal rotation (Varimax) only. 2 - Perform an oblique rotation (Promax, including Varimax).
33-34	Constant for number of factors to rotate 0 - Rotate the number of factors determined by the option chosen in cc 25-26 above. k - Rotate a number of factors equal to the minimum of k, ten, and/or the number of factors determined by the option above in cc 25-26.
35-36	Pooling option (see Sections 2.5.3 and 2.2.4) 00 - Add matrices with ID = 1 Nonzero - Subtract matrices with ID = 1

Note: In columns 37-60, the matrix output options are as follows:

0 - No output
1 - Print only
2 - Print and punch
3 - Punch

Column	Meaning
37-38	Output the latent vectors, A*
39-40	Output the unrotated factor matrix, F
41-42	Output the orthogonal transformation matrix, T
43-44	Output the orthogonal factor matrix, G
45-46	Output the transformation matrix to oblique vector structure, L
47-48	Output the oblique reference vector structure matrix, V
49-50	Output the correlations among oblique reference vectors, ψ
51-52	Output the oblique reference vector pattern matrix, W
53-54	Output the correlations between reference vectors and primary factors, D
55-56	Output the oblique primary factor structure matrix, S
57-58	Output the correlations among oblique primary factors, Φ
59-60	Output the oblique primary factor pattern matrix, P
61-62	Output the factor score regression coefficients, β

VARIABLE NAME CARD

In the factor analysis program there are a number of matrix printouts that the user may request. The variables in the matrix may be assigned a four-character name to aid in the identification of the output. The card is punched in four-column fields, and each field corresponds to the variable to be identified (for example, field 3 (columns 9-12) will be the name of row and column 3 on all matrix output). At most, 20 names can appear on one card. If there are more than 20 variables in the analysis, a second card having the same format as the first must be included in the control card deck.

Column	Meaning
1-4	Name of variable 1.
5-8	Name of variable 2.
....	....
(4N-3) - (4N)	Name of variable N.

2.2.5 Data Input

Raw data input to the program consists of a set of observations made on several different variables. The variables for each observation are punched on one, two, or three cards, according to the following general format:

Field	Type	Meaning
1	Integer (I)	Identification field. Any numeric information that serves to identify the particular observation is punched in this field. It must be greater than zero, and should be different for each observation.
2	Integer (I)	Card number within observation. If it is not possible for one card to contain all the variables, they may be continued on a second and a third card, as necessary. The user has the option of sequence checking the cards to ensure that all cards within a case are together, and that the order of cards is consistent. If the option is chosen (cc 5-6 on option card), this field must be punched with an integer that is in ascending sequence for all cards in the case. If sequence checking is not desired, the field may be blank and may consist of one blank column.
3,4....,n etc.	Floating point (F)	Variable x_1. Any number may be punched in this field. Decimal points are not required. The remaining fields on the card are reserved for variables. If there are more variables than can fit on the first card, a second and a third card may be used.

Following the data deck, the user must include a card containing a negative integer in the identification field. This card signals the end of data.

2.2.6 Matrix Input/Output

It is possible to obtain punched card output of a number of matrices (see section 2.5.3) and vectors with this program. This program is designed to also input some of these matrices, at a later time, for further analysis or processing. In addition, matrices from another program or source, if punched in the program format, may also be used as input.

This section is devoted to a description of various possible forms of analysis with the output options available in each program.

Format Description

See the matrix format given under "Format Description" in section 2.1.4.

Factor Analysis with Correlation Matrix Input

The punched output option of the correlation matrix includes the punchout of the number of cases (matrix 21), and means and standard deviation vectors (matrix 23). This complete output can be used as input to initiate another analysis without the necessity of reprocessing the source data that was used to generate the matrices.

To use the correlation matrix set as input, the user places the punched output behind the variable names card, followed by a blank card, or one that contains a negative number in the job number field. The program reads the number of cases, means, and standard deviations and correlation matrix, stores each in its appropriate location, and then initiates the analysis as specified on the option card.

Pooling Sums of Squares and Cross Products (cc 35-36)

The topic is discussed fully in section 2.1.4. When raw sums of squares matrices have been previously punched by this program (or by hand), in accordance with the above format description, they can be stacked and will be combined by using this option. When the option is given the number 0, they will be added. If a nonzero field is used, all matrices will be added until the first negative (left-justified) job number field (cc 1-4) is encountered. Subsequent matrices will be subtracted until the second negative problem number card is encountered.

Iterating on Communalities*

This program does not iterate on communalities, which is a desirable feature mentioned in section 2.2. However, by electing to punch the correlation matrix, one can insert the estimated communalities on the diagonal of this matrix and iterate using matrix input.

*The factor scores computation reads the original data matrix, X, from the disk. If X is not on the disk, which is the case if any card reader input with other data has been read subsequent to the reading of X, then X must be reread under input mode 1 before this matrix input option is used.

2.2.7 Operating Instructions

A. Using the factor analysis program when the total 1130 Statistical System has not been stored on the disk

If the user wishes to load only the set of programs that allow this type of analysis, the following programs must be compiled or assembled and stored on the disk. Each deck begins with a card punched as

// FOR

and ends with an

*STORE

card.

The user should use a disk containing the 1130 Disk Monitor System, as described in section 1.1. The following decks should be preceded by a cold start card, placed in the card reader hopper, and the buttons IMMEDIATE STOP (console), RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. A blank card should be placed after the last deck in the card reader hopper.

> DECKS-LABELS: FCTR-FCTR; FCTR1-FCT1; FCTR2-FCT2; FCTR3-FCT3; *FMTRD-FMRD; *DATRD-DTRD; *GMPYX-GMPY; *GDIVX-GDIV; *PRNTB-PRNB; **COREL-CORL; **PRNT-PRNT; **MXRAD-MXRD; INVRS-INVS; XMAX-XMAX; TRIDI-TRID; QR-QR; VECTR-VCTR; COVEC-CVEC; RFOUT-ROUT; PROMX-PRMX; VARMX-VRMX; RPRNT-RPNT; MATIN-MATN; SCORE-SCOR; *FMAT-FMAT; TRAN-TRAN.

In addition, regression and factor analysis programs must reside on the disk together; section 2.1.5 names additional routines to be placed on the disk.

B. Execution from disk

Once the component subroutines and main calling programs are on the disk, the execution of a job requires the monitor control cards, program control cards, and data cards to be placed in the card reader. The deck should be preoeded by a cold start card. To initiate processing, the buttons IMMEDIATE STOP and RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. The order in which the cards are placed in the card reader for either matrix or raw data input is shown in Figures 8, 9, and 10.

*Used in all four analysis types
**Used in regression analysis

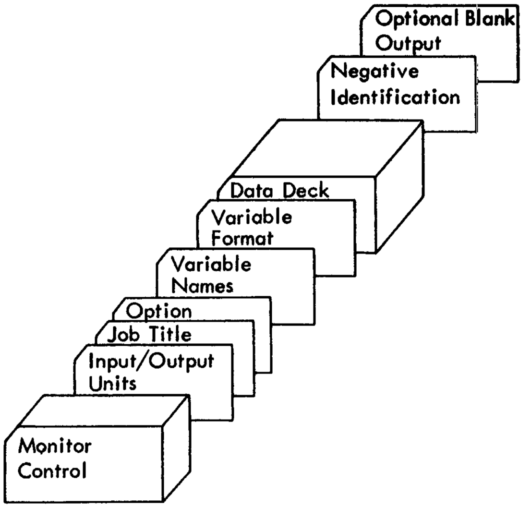


Figure 8. Factor analysis card order — card reader input

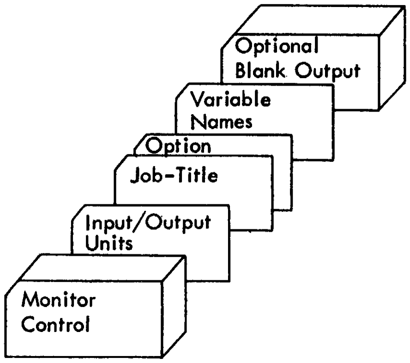


Figure 9. Factor analysis card order — disk input

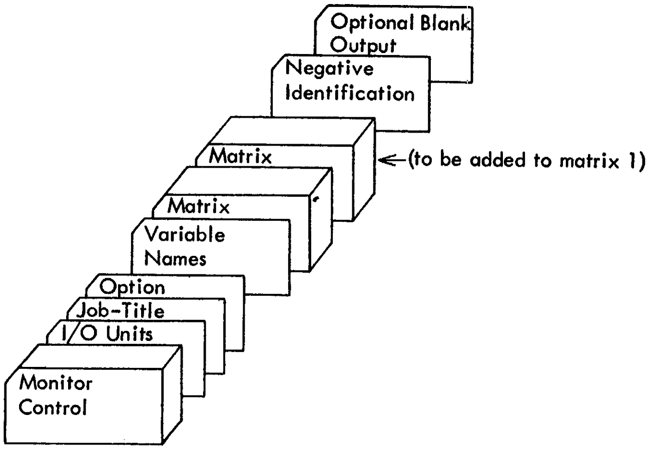


Figure 10. Factor analysis card order — matrix input

2.2.8 Sample Problem

```
                      INPUT

// XEQ FCTR     05
*LOCALFCTR,FMTRD,DATRD,PRNTB,MXRAD,TRAN
*LOCALFCTR1,TRIDI,QR,INVRS
*LOCALCOREL,PRNT
*LOCALFCTR2,VECTR,PRNT
*LOCALFCTR3,VARMX,PROMX,SCORE,RFOUT
020200
3333    FACTOR ANALYSIS SAMPLE PROBLEM
040100000000000101010202020200020000 1010101010101010101010101
  P1  P2  P3  P4
I2I2,1X,4F6.0)
0101 000063000075000159000041
0201 000101000092000142000049
0301 000119000098000131000068
0401 000157000101000124000092
0501 000178000104000119000097
0601 000147000106000118000102
0701 000128000108000116000109
0801 000113000107000116000066
0901 000094000107000115000044
1001 000111000104000117000069
1101 000139000110000104000117
1201 000157000107000100000118
1301 000169000111000075000157
1401 000145000109000079000107
1501 000079000095000096000069
1601 000049000086000111000047
1701 000048000077000111000032
1801 000041000069000106000022
1901 000066000062000097000017
2001 000111000074000092000045
2101 000164000104000088000097
2201 000170000117000039000164
2301 000208000135000053000246
2401 000237000148000058000366
2501 000169000152000061000230
2601 000114000137000073000175
2701 000106000130000077000178
2801 000097000123000086000156
2901 000099000110000092000125
3001 000111000111000102000105
3101 000068000108000108000081
3201 000048000096000121000044
3301 000042000078000123000020
3401 000034000073000125000017
3501 000048000084000125000014
-1

      PUNCHED CORRELATION MATRIX OUTPUT

3333 4 1 1 0.1000000E 01 0.7225732E 00-0.5798441E 00 0.7999757E 00
3333 4 1 2 0.7225732E 00 0.1000000E 01-0.6567597E 00 0.8950214E 00
3333 4 1 3-0.5798441E 00-0.6567597E 00 0.1000000E 01-0.7525222E 00
3333 4 1 4 0.7999757E 00 0.8950214E 00-0.7525222E 00 0.1000000E 01
333323 1 1 0.1122857E 03 0.5141101E 02
333323 1 2 0.1030857E 03 0.2161068E 02
333323 1 3 0.1016857E 03 0.2601331E 02
333323 1 4 0.9960000E 02 0.7545321E 02
333321 1 1 0.3500000E 02
```

```
     PUNCHED FACTOR SCORES OUTPUT

  125 0.1952047E 01-0.2331642E 01
  225 0.2963847E 00-0.1594353E 01
  325-0.4612540E 00-0.1131953E 01
  425-0.2389709E 01-0.7728281E 00
  525-0.3506941E 01-0.5253528E 00
  625-0.1578879E 01-0.5796633E 00
  725-0.3429788E 00-0.5622065E 00
  825 0.6738971E-02-0.5635983E 00
  925 0.8490992E 00-0.5561192E 00
 1025 0.3999518E-01-0.6032414E 00
 1125-0.1011253E 01-0.5682308E-01
 1225-0.2248286E 01 0.1632854E 00
 1325-0.2765937E 01 0.1167846E 01
 1425-0.1946555E 01 0.9881128E 00
 1525 0.1136240E 01 0.1907818E 00
 1625 0.2493227E 01-0.4555233E 00
 1725 0.1984453E 01-0.4183342E 00
 1825 0.1837948E 01-0.2014065E 00
 1925-0.1365285E 00 0.2584076E 00
 2025-0.1992324E 01 0.5247329E 00
 2125-0.3228456E 01 0.6974968E 00
 2225-0.3100469E 01 0.2611022E 01
 2325-0.3355663E 01 0.2021247E 01
 2425-0.3044062E 01 0.1755776E 01
 2525-0.3732885E 00 0.1551542E 01
 2625 0.1751099E 01 0.1005148E 01
 2725 0.2006701E 01 0.8391422E 00
 2825 0.2129229E 01 0.4882641E 00
 2925 0.1194178E 01 0.3152475E 00
 3025 0.4921422E 00-0.4477804E-01
 3125 0.2693805E 01-0.3768844E 00
 3225 0.3135445E 01-0.8926919E 00
 3325 0.2451350E 01-0.9157080E 00
 3425 0.2694848E 01-0.1001286E 01
 3525 0.2337661E 01-0.9936804E 00

                    OUTPUT
```

```
// XEQ FCTR     05

*LOCALFCTR,FMTRD,DATRD,PRNTB,MXRAD,TRAN
*LOCALFCTR1,TRIDI,QR,INVRS
*LOCALCOREL,PRNT
*LOCALFCTR2,VECTR,PRNT
*LOCALFCTR3,VARMX,PROMX,SCORE,RFOUT
```

```
         FACTOR ANALYSIS SAMPLE PROBLEM                                                 JOB   3333      PAGE       0

          NUMBER OF VARIABLES                                                  4

          INPUT TYPE                                                           1
          SEQUENCE CHECK                                                       0
          VARIABLES ON CARD 1                                                  0
          VARIABLES ON CARD 2                                                  0
          VARIABLES ON CARD 3                                                  0
          TRANSFORMATION SWITCH                                                0
          OUTPUT RAW CROSS PRODUCTS                                            1

          OUTPUT RESIDUAL CROSS PRODUCTS                                       1
          OUTPUT VARIANCE - COVARIANCE                                         1
          OUTPUT CORRELATION                                                   2
          FACTOR SCORES                                                        2
          NUMBER OF FACTORS OPTION                                             2
          NUMBER OF FACTORS OR PERCENT OF TRACE                                2

          COMMUNALITY OPTION                                                   0
          ROTATION OPTION                                                      2
          NUMBER OF FACTORS TO ROTATE                                          0
          POOLING OPTION                                                       0
          LATENT VECTORS                                                       1
          UNROTATED FACTOR MATRIX                                              1
          ORTHOGONAL TRANSFORMATION MATRIX                                     1

          ORTHOGONAL FACTOR MATRIX                                             1
          TRANSFORMATION MATRIX TO OBLIQUE REFERENCE VECTOR STRUCTURE          1

          OBLIQUE REFERENCE VECTOR STRUCTURE MATRIX                            1
          CORRELATIONS AMONG OBLIQUE REFERENCE VECTORS                         1
          OBLIQUE REFERENCE VECTOR PATTERN MATRIX                              1

          CORRELATIONS BETWEEN REFERENCE VECTORS AND PRIMARY FACTORS           1

          OBLIQUE PRIMARY FACTOR STRUCTURE MATRIX                              1
          CORRELATIONS AMONG OBLIQUE PRIMARY FACTORS                           1
          OBLIQUE PRIMARY FACTOR PATTERN MATRIX                                1
          FACTOR SCORE REGRESSION COEFFICIENTS                                 1

(2I2,1X,4F6.0)
```

```
        FACTOR ANALYSIS SAMPLE PROBLEM                                                  JOB   3333      PAGE       1

                                                  MATRIX OF RAW CROSS-PRODUCTS


VARIABLE        P1             P2             P3             P4
     P1    0.53114E 06   0.43242E 06   0.37325E 06   0.49693E 06
     P2    0.43242E 06   0.38781E 06   0.35432E 06   0.40897E 06
     P3    0.37325E 06   0.35432E 06   0.38490E 06   0.30425E 06

     P4    0.49693E 06   0.40897E 06   0.30425E 06   0.54077E 06
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE      2

                                             MATRIX OF RESIDUAL CROSS-PRODUCTS


VARIABLE          P1           P2           P3           P4
   P1       0.89865E 05  0.27295E 05 -0.26365E 05  0.10550E 06

   P2       0.27295E 05  0.15878E 05 -0.12553E 05  0.49620E 05
   P3      -0.26365E 05 -0.12553E 05  0.23007E 05 -0.50219E 05
   P4       0.10550E 06  0.49620E 05 -0.50219E 05  0.19356E 06
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE      3
                                               VARIANCE - COVARIANCE MATRIX


VARIABLE          P1           P2           P3           P4
   P1       0.26430E 04  0.80279E 03 -0.77546E 03  0.31032E 04

   P2       0.80279E 03  0.46702E 03 -0.36920E 03  0.14594E 04
   P3      -0.77546E 03 -0.36920E 03  0.67669E 03 -0.14770E 04
   P4       0.31032E 04  0.14594E 04 -0.14770E 04  0.56931E 04
```

```
FACTOR ANALYSIS SAMPLE PROBLEM                                                    JOB    3333      PAGE      4


 SUMMARY STATISTICS           NO.OF CASES=    35
    VARIABLE                    LOW          HIGH         AVERAGE      STD. DEV.    VARIANCE


     1     P1             0.34000E 02  0.23700E 03  0.11228E 03  0.51411E 02  0.26430E 04
     2     P2             0.62000E 02  0.15200E 03  0.10308E 03  0.21610E 02  0.46702E 03
     3     P3             0.39000E 02  0.15900E 03  0.10168E 03  0.26013E 02  0.67669E 03
     4     P4             0.14000E 02  0.36600E 03  0.99600E 02  0.75453E 02  0.56931E 04
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE      5
                                             MATRIX OF CORRELATION COEFFICIENTS


VARIABLE          P1           P2           P3           P4

   P1       0.10000E 01  0.72257E 00 -0.57984E 00  0.79997E 00
   P2       0.72257E 00  0.10000E 01 -0.65675E 00  0.89502E 00
   P3      -0.57984E 00 -0.65675E 00  0.10000E 01 -0.75252E 00
   P4       0.79997E 00  0.89502E 00 -0.75252E 00  0.10000E 01
   FACTOR ANALYSIS SAMPLE PROBLEM                                                       JOB    3333       PAGE       6
                                                MATRIX OF CHARACTERISTIC VECTORS
P1     -0.48311E 00 -0.54530E 00
P2     -0.51266E 00 -0.16668E 00
P3      0.46188E 00 -0.81965E 00
P4     -0.53892E 00 -0.55082E-01
```

```
FACTOR ANALYSIS SAMPLE PROBLEM                                                          JOB    3333       PAGE       7

                                    TRACE          4.0000

                                  CHARACTERISTIC ROOTS           CUMUL. PERCENT OF TRACE


                                         3.2134                        80.3374
                                         0.4301                        91.0900
                                         0.2753                         0.0000
                                         0.0810                         0.0000
```

```
   FACTOR ANALYSIS SAMPLE PROBLEM                                                       JOB    3333       PAGE       8
                                             NORMALIZED UNROTATED FACTOR LOADINGS
P1     -0.86604E 00 -0.35762E 00
P2     -0.91901E 00 -0.10931E 00
P3      0.82798E 00 -0.53754E 00
P4     -0.96608E 00 -0.36124E-01

   COMMUNALITIES

      0.8779364E 00
      0.8565298E 00


      0.9745132E 00
      0.9346225E 00
```

```
FACTOR ANALYSIS SAMPLE PROBLEM                                                          JOB    3333       PAGE       9


                                     NORMAL VARIMAX CRITERION (NORMALIZED)

                          CYCLE       CRITERION      DIFFERENCE     EPSILON CRITERION=      0.00116000

                             1       0.02850188      0.02850188
                             2       0.18610206      0.15760016
                             3       0.18610206      0.00000000
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                                   JOB    3333      PAGE      10
                                                ORTHOGONAL TRANSFORMATION MATRIX
VARIABLE              1             2
   1               0.7966       -0.6044
   2               0.6044        0.7966
```

```
       FACTOR ANALYSIS SAMPLE PROBLEM                                               JOB   3333     PAGE    11
                                                  ORTHOGONAL FACTOR MATRIX(VARIMAX)
VARIABLE          1          2

   P1         -0.9061     0.2385
   P2         -0.7982     0.4684
   P3          0.3346    -0.9287
   P4         -0.7914     0.5551
```

```
       FACTOR ANALYSIS SAMPLE PROBLEM                                               JOB   3333     PAGE    12

                                                     TRANSFORMATION TO OBLIQUE REFERENCE VECTOR STRCTR.

VARIABLE          1          2
  1           0.9322     0.3853
  2           0.3617     0.9227
```

```
       FACTOR ANALYSIS SAMPLE PROBLEM                                               JOB   3333     PAGE    13
                                                     CORRELATIONS AMONG OBLIQUE REFERENCE VECTORS

VARIABLE          1          2
  1           1.0000     0.6931
  2           0.6931     1.0000
```

```
       FACTOR ANALYSIS SAMPLE PROBLEM                                               JOB   3333     PAGE    14

                                                     OBLIQUE REFERENCE VECTOR STRUCTURE MATRIX

VARIABLE          1          2

   P1         -0.7584    -0.1290

   P2         -0.5746     0.1246
   P3         -0.0239    -0.7279
   P4         -0.5370     0.2072
```

```
       FACTOR ANALYSIS SAMPLE PROBLEM                                               JOB   3333     PAGE    15
                                                     OBLIQUE REFERENCE VECTOR PATTERN MATRIX

VARIABLE          1          2
   P1         -1.2874     0.7632
   P2         -1.2722     1.0063
   P3          0.9249    -1.3690
   P4         -1.3099     1.1151
```

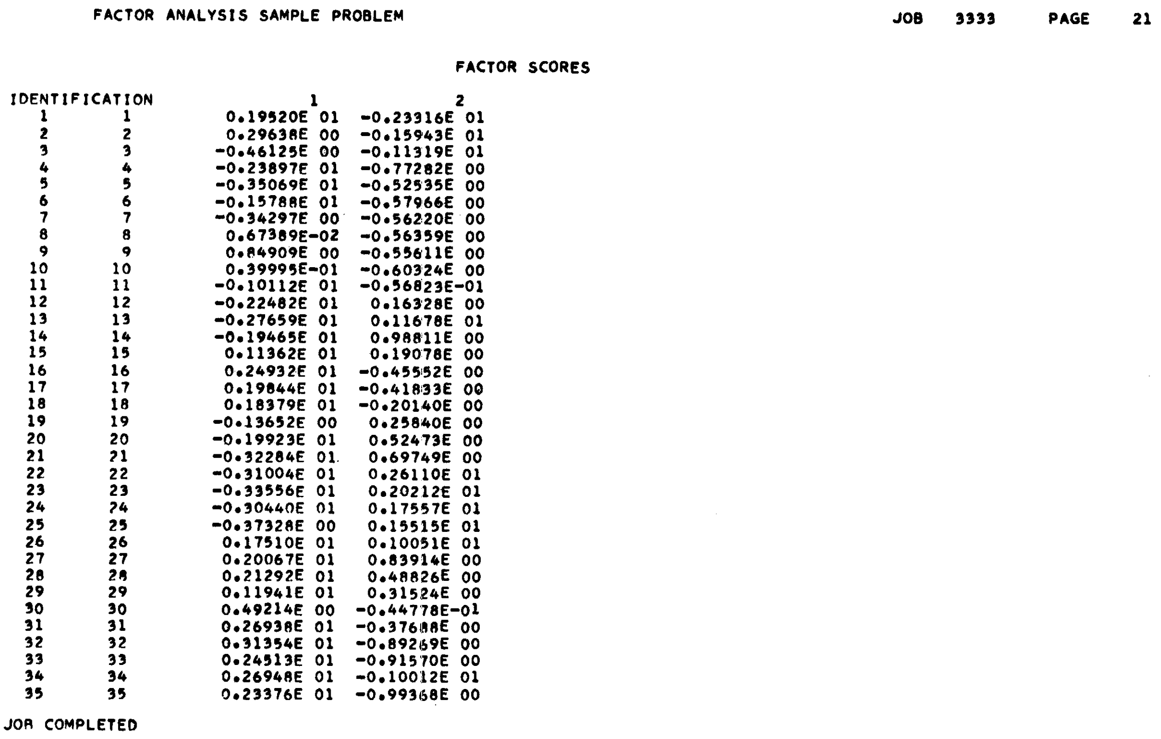
```
          FACTOR ANALYSIS SAMPLE PROBLEM                                                      JOB   3333     PAGE    16

                                                       CORR. BET. REFERENCE VECTORS AND PRIMARY FACTORS

VARIABLE              1           2
   1               0.7208      0.0000
   2               0.0000      0.7208
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                                      JOB   3333     PAGE    17

                                                       CORR. AMONG OBLIQUE PRIMARY FACTORS
VARIABLE              1           2
   1               1.0000     -0.6931
   2              -0.6931      1.0000
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                                      JOB   3333     PAGE    18

                                                       OBLIQUE PRIMARY FACTOR STRUCTURE MATRIX
VARIABLE              1           2
     P1           -0.9280      0.5502

     P2           -0.9170      0.7254

     P3            0.6667     -0.9868

     P4           -0.9442      0.8038
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                                      JOB   3333     PAGE    19

                                                       OBLIQUE PRIMARY FACTOR LOADINGS
VARIABLE              1           2
     P1           -1.0521     -0.1790
     P2           -0.7972      0.1728
     P3           -0.0332     -1.0098
     P4           -0.7450      0.2875
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                                      JOB   3333     PAGE    20

                                                       FACTOR SCORE REGRESSION COEFFICIENTS
VARIABLE              1           2
     P1           -2.9865      0.1404
     P2            0.9614     -0.0652
     P3            0.4487     -1.0582
     P4            0.8373     -0.0641
```



```
     FACTOR ANALYSIS SAMPLE PROBLEM                                                   JOB   3333      PAGE     21

                                                  FACTOR SCORES

IDENTIFICATION             1           2
   1        1         0.19520E 01  -0.23316E 01
   2        2         0.29638E 00  -0.15943E 01
   3        3        -0.46125E 00  -0.11319E 01
   4        4        -0.23897E 01  -0.77282E 00
   5        5        -0.35069E 01  -0.52535E 00
   6        6        -0.15788E 01  -0.57966E 00
   7        7        -0.34297E 00  -0.56220E 00
   8        8         0.67389E-02  -0.56359E 00
   9        9         0.84909E 00  -0.55611E 00
  10       10         0.39995E-01  -0.60324E 00
  11       11        -0.10112E 01  -0.56823E-01
  12       12        -0.22482E 01   0.16328E 00
  13       13        -0.27659E 01   0.11678E 01
  14       14        -0.19465E 01   0.98811E 00
  15       15         0.11362E 01   0.19078E 00
  16       16         0.24932E 01  -0.45552E 00
  17       17         0.19844E 01  -0.41833E 00
  18       18         0.18379E 01  -0.20140E 00
  19       19        -0.13652E 00   0.25840E 00
  20       20        -0.19923E 01   0.52473E 00
  21       21        -0.32284E 01   0.69749E 00
  22       22        -0.31004E 01   0.26110E 01
  23       23        -0.33556E 01   0.20212E 01
  24       24        -0.30440E 01   0.17557E 01
  25       25        -0.37328E 00   0.15515E 01
  26       26         0.17510E 01   0.10051E 01
  27       27         0.20067E 01   0.83914E 00
  28       28         0.21292E 01   0.48826E 00
  29       29         0.11941E 01   0.31524E 00
  30       30         0.49214E 00  -0.44778E-01
  31       31         0.26938E 01  -0.37688E 00
  32       32         0.31354E 01  -0.89269E 00
  33       33         0.24513E 01  -0.91570E 00
  34       34         0.26948E 01  -0.10012E 01
  35       35         0.23376E 01  -0.99368E 00

JOB COMPLETED
```

```
CORRELATION MATRIX INPUT--MULTIPLE R**2 ON DIAGONAL

// XEQ FCTR     05
*LOCALFCTR,FMTRD,DATRD,PRNTB,MXRAD,TRAN
*LOCALFCTR1,TRIDI,QR,INVRS
*LOCALCOREL,PRNT
*LOCALFCTR2,VECTR,PRNT
*LOCALFCTR3,VARMX,PROMX,SCORE,RFOUT
020200
3333    FACTOR ANALYSIS SAMPLE PROBLEM
04030000000000000000100020200020000 10100010101010001010100
  P1   P2   P3   P4
3333 4 1 1 0.6412571E 00 0.7225732E 00-0.5798441E 00 0.7999757E 00
3333 4 1 2 0.7225732E 00 0.8018028E 00-0.6567597E 00 0.8950214E 00
3333 4 1 3-0.5798441E 00-0.6567597E 00 0.5689992E 00-0.7525222E 00
3333 4 1 4 0.7999757E 00 0.8950214E 00-0.7525222E 00 0.8816457E 00
333323 1 1 0.1122857E 03 0.5141101E 02
333323 1 2 0.1030857E 03 0.2161068E 02
333323 1 3 0.1016857E 03 0.2601331E 02
333323 1 4 0.9960000E 02 0.7545321E 02
333321 1 1 0.3500000E 02
  -1

                         OUTPUT
```

```
// XEQ FCTR      05

*LOCALFCTR,FMTRD,DATRD,PRNTB,MXRAD,TRAN
*LOCALFCTR1,TRIDI,QR,INVRS
*LOCALCOREL,PRNT
*LOCALFCTR2,VECTR,PRNT
*LOCALFCTR3,VARMX,PROMX,SCORE,RFOUT
```

```
                                                                                      JOB    3333      PAGE      0
FACTOR ANALYSIS SAMPLE PROBLEM

 NUMBER OF VARIABLES                                                    4
 INPUT TYPE                                                             3
 SEQUENCE CHECK                                                         0
 VARIABLES ON CARD 1                                                    0
 VARIABLES ON CARD 2                                                    0
 VARIABLES ON CARD 3                                                    0
 TRANSFORMATION SWITCH                                                  0
 OUTPUT RAW CROSS PRODUCTS                                              0
 OUTPUT RESIDUAL CROSS PRODUCTS                                         0
 OUTPUT VARIANCE - COVARIANCE                                           0
 OUTPUT CORRELATION                                                     1
 FACTOR SCORES                                                          0
 NUMBER OF FACTORS OPTION                                               2
 NUMBER OF FACTORS OR PERCENT OF TRACE                                  2
 COMMUNALITY OPTION                                                     0
 ROTATION OPTION                                                        2
 NUMBER OF FACTORS TO ROTATE                                            0

 POOLING OPTION                                                         0
 LATENT VECTORS                                                         1
 UNROTATED FACTOR MATRIX                                                1
 ORTHOGONAL TRANSFORMATION MATRIX                                       0
 ORTHOGONAL FACTOR MATRIX                                               1
 TRANSFORMATION MATRIX TO OBLIQUE REFERENCE VECTOR STRUCTURE            1

 OBLIQUE REFERENCE VECTOR STRUCTURE MATRIX                              1
 CORRELATIONS AMONG OBLIQUE REFERENCE VECTORS                           1
 OBLIQUE REFERENCE VECTOR PATTERN MATRIX                                0
 CORRELATIONS BETWEEN REFERENCE VECTORS AND PRIMARY FACTORS             1
 OBLIQUE PRIMARY FACTOR STRUCTURE MATRIX                                1

 CORRELATIONS AMONG OBLIQUE PRIMARY FACTORS                             1
 OBLIQUE PRIMARY FACTOR PATTERN MATRIX                                  1
 FACTOR SCORE REGRESSION COEFFICIENTS                                   0
```

```
                                                                                      JOB    3333      PAGE      1
   FACTOR ANALYSIS SAMPLE PROBLEM
                                              MATRIX OF CHARACTERISTIC VECTORS
P1      -0.46728E 00 -0.47333E 00
P2      -0.52362E 00 -0.31961E 00
P3       0.43527E 00 -0.81887E 00
P4      -0.56391E 00  0.56930E-01
```

```
                                                                                      JOB    3333      PAGE      2
FACTOR ANALYSIS SAMPLE PROBLEM

                                   TRACE          2.8937

                                 CHARACTERISTIC ROOTS          CUMUL. PERCENT OF TRACE


                                         2.9564                    102.1689

                                         0.0298                    103.1991
                                        -0.0061                      0.0000
                                        -0.0864                      0.0000
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                                          JOB    3333       PAGE       3
                                                 NORMALIZED UNROTATED FACTOR LOADINGS
P1      -0.80346E 00 -0.81725E-01
P2      -0.90033E 00 -0.55183E-01
P3       0.74842E 00 -0.14138E 00
P4      -0.96961E 00  0.98294E-02

      COMMUNALITIES

          0.6522407E 00
          0.8136515E 00
          0.5801334E 00

          0.9402521E 00
```

```
FACTOR ANALYSIS SAMPLE PROBLEM                                                                JOB    3333       PAGE       4

                                             NORMAL VARIMAX CRITERION (NORMALIZED)

                        CYCLE       CRITERION      DIFFERENCE     EPSILON CRITERION=      0.00116000
                          1        0.00035874     0.00035874

                          2        0.02373140     0.02337265

                          3        0.02373140     0.00000000
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                                         JOB    3333       PAGE       5
                                                    ORTHOGONAL FACTOR MATRIX(VARIMAX)
VARIABLE               1            2
     P1            -0.6504       0.4786

     P2            -0.7044       0.5633

     P3             0.4599      -0.6071
     P4            -0.7121       0.6580
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                                        JOB    3333       PAGE       6
                                                       TRANSFORMATION TO OBLIQUE REFERENCE VECTOR STRCTR.
VARIABLE               1            2
    1               0.8730       0.3683
    2               0.4876       0.9296
```

```
      FACTOR ANALYSIS SAMPLE PROBLEM                                                        JOB    3333       PAGE       7


                                                       CORRELATIONS AMONG OBLIQUE REFERENCE VECTORS
VARIABLE               1            2
    1               1.0000       0.7749
    2               0.7749       1.0000
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE      8
                                                      OBLIQUE REFERENCE VECTOR STRUCTURE MATRIX
VARIABLE           1          2
     P1         -0.3344     0.2054
     P2         -0.3403     0.2642

     P3          0.1054    -0.3950
     P4         -0.3008     0.3494
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE      9
                                                      CORR. BET. REFERENCE VECTORS AND PRIMARY FACTORS


VARIABLE           1          2
   1             0.6320     0.0000
   2             0.0000     0.6320
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE     10
                                                      CORR. AMONG OBLIQUE PRIMARY FACTORS

VARIABLE           1          2
   1             1.0000    -0.7749
   2            -0.7749     1.0000
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE     11

                                                      OBLIQUE PRIMARY FACTOR STRUCTURE MATRIX

VARIABLE           1          2

     P1         -0.7810     0.7351

     P2         -0.8624     0.8353

     P3          0.6512    -0.7543
     P4         -0.9045     0.9218
```

```
          FACTOR ANALYSIS SAMPLE PROBLEM                                              JOB    3333      PAGE     12

                                                      OBLIQUE PRIMARY FACTOR LOADINGS

VARIABLE           1          2
     P1         -0.5291     0.3250

     P2         -0.5384     0.4181

     P3          0.1668    -0.6250

     P4         -0.4760     0.5529

JOB COMPLETED
```

2.2.9 References

Businger, P.A. ''Eigenvalues of a real symmetric matrix by the QR method'', Algorithm 253, Communications of the Association for Computing Machinery, April 1965, vol. 8, no. 4.

Guttman, L. ''Some Necessary Conditions for Common Factor Analysis'', Psychometrika, 1954, 19, 149-162.

Harman, H. Modern Factor Analysis. University of Chicago Press, 1960, pp. 289-308; pp. 261-288; pp. 349-356.

Hendrickson and White. ''Promax: A quick method for rotation to oblique simple structure'', British Journal of Statistical Psychology, 1964, XVII, 65-70.

Horst, P. Factor Analysis of Data Matrices. Holt, Rinehart, and Winston, Inc. New York, Chicago, San Francisco, Toronto, London. 1965, p. 214.

HOW. FORTRAN Subroutine, using non-iterative methods of Householder, Ortega, and Wilkinson, solves for eigenvalues and corresponding eigenvectors of a real symmetric matrix. Program Writeup F2 BC HOW. Berkeley Division, University of California, 1962.

Kaiser, ''The Varimax criterion for analytical rotation in factor analysis'', Psychometrika, 1958, 23, 187-200.

Wilkinson, J.H. ''The calculation of the eigenvectors of codiagonal matrices'', The Computer Journal, 1958, vol. 1, p. 90.

Wilkinson, J.H. "Householder's method for the solution of the algebraic eigen problem", The Computer Journal, 1960, vol. 3, p. 23.

2.3 ANALYSIS OF VARIANCE

From the experimental observations on a variable x, this program will compute an analysis of variance for a complete factorial design for a maximum of four (4) factors. The method used in the program is essentially that described by H.O. Hartley.* This method is particularly useful, since it can be extended to accommodate a great many experimental designs.

The extension to other experimental designs is accomplished by a very simple procedure. The program performs a factorial analysis and then allows the user to pool certain components of the analysis of variance table in accordance with the summary instructions that specifically apply to the particular design desired. For example, a two- or three-factor design can result in the following analysis of variance tables:

- Single classification
- Two-way classification with cell repetition
- Randomized block with two factor treatments
- Split plot
- Split-split plot
- Three-factor randomized blocks

By utilizing a special report generator, the user has flexibility in choosing the appropriate components to pool in forming the error term or terms to accommodate the above designs or any other similar designs.

Once the data is contained in storage, the sum of squares is computed as follows:

Let A_i = Sum of all the observations at level a_i

n_a = Number of observations summed to obtain A_i

AB_{ij} = Sum of all observations at level ab_{ij}

n_{ab} = Number of observations summed to obtain AB_{ij}

ABC_{ijk} = Sum of all observations at level abc_{ijk}

n_{abc} = Number of observations summed to obtain ABC_{ijk}

Thus, a general formula for the main effect due to factor A is:

$$SS_a = \frac{\Sigma A^2_i}{n_a} - \frac{G^2}{n_g}$$

*Ralston, A. and Wilf, H. S., Mathematical Methods for Digital Computers. New York: John Wiley and Sons, Inc., 1960, pp. 221-230.

where G = the grand total of all observations, and n_g is the number of observations summed to obtain G. The main effect due to factor B has the form:

$$SS_b = \frac{\Sigma B^2_i}{n_b} - \frac{G^2}{n_g}$$

The general computational formula for the variation due to the AB interaction is:

$$SS_{ab} = \frac{\Sigma (AB_{ij})^2}{n_{ab}} - \frac{G^2}{n_g} - (SS_a + SS_b)$$

The general computational formula for the variation due to the ABC interaction is:

$$SS_{abc} = \frac{\Sigma (ABC_{ijk})^2}{n_{abc}} - \frac{G^2}{n_g} - (SS_a + SS_b + SS_c + SS_{ab} + SS_{ac} + SS_{bc})$$

The following table shows the layout of the complete table for the analysis of variance of a complete factorial design:

ANOVA TABLE FOR A COMPLETE FACTORIAL DESIGN

Source of Variation	D.F.	Sum of Squares	Mean Square
A Main effect	(p-1)	SS_a	$SS_a/$ (p-1)
B Main effect	(q-1)	SS_b	$SS_b/$ (q-1)
C Main effect	(r-1)	SS_c	$SS_c/$ (r-1)
AB Interaction	(p-1)(q-1)	SS_{ab}	$SS_{ab}/$ (p-1)(q-1)
AC Interaction	(p-1)(r-1)	SS_{ac}	$SS_{ac}/$ (p-1)(r-1)
BC Interaction	(q-1)(r-1)	SS_{bc}	$SS_{bc}/$ (q-1)(r-1)
ABC Interaction	(p-1)(q-1)(r-1)	SS_{abc}	$SS_{abc}/$ (p-1)(q-1)(r-1)
Experimental error (within cell)	pqr (n-1)	SS_{error}	$SS_{error}/$ pqr (n-1)
Total	npqr-1	SS_{total}	

where p = number of levels in factor A
q = number of levels in factor B
r = number of levels in factor C
n = number of observations per cell

To obtain other experimental designs from a complete factorial design, the user should analyze the data as if it were a complete factorial design, and then reconstruct his ANOVA table from the output.

Not all experimental designs can be handled by this technique, notably Latin and Youden squares, lattices, and incomplete randomized blocks.

Also, this program does not handle repeated measurement designs (that is, replications must be considered as a factor). For a detailed account of the various experimental designs, see O. Kempthorne, Design and Analysis of Experiments (John Wiley, 1952).

Single Classification Design (A X B). In this case, the replications are considered as a factor (B). The error term is:

$$SS_{error} = SS_b + SS_{ab}$$

giving the following reconstructed ANOVA table:

$$\begin{array}{l} SS_a \\ SS_{error} \\ \hline SS_{total} \end{array}$$

Two-Way Classification with Cell Repetition (A X B X C). This differs from a randomized block design in that one is not interested in the recovery of interblock information. Consequently, the error term is:

$$SS_{error} = SS_c + SS_{ac} + SS_{bc} + SS_{abc}$$

(where factor C is the cell repetitions) thus giving the following ANOVA table:

$$\begin{array}{l} SS_a \\ SS_b \\ SS_{ab} \\ SS_{error} \\ \hline SS_{total} \end{array}$$

Randomized Block with Two Treatment Factors (A X B X C). Here the third "factor" (C) is blocks. In this case, one is interested in finding out whether there are significant differences between blocks, so the error term is computed from:

$$SS_{error} = SS_{ac} + SS_{bc} + SS_{abc}$$

thus giving an ANOVA table as follows:

$$\begin{array}{l} SS_a \\ SS_b \\ SS_{ab} \\ SS_c \text{ (blocks)} \\ SS_{error} \\ \hline SS_{total} \end{array}$$

Split-Plot Design (A X B X C). In this case, let factor A = main treatments; B = subtreatments, and C = blocks. Then, appropriate error terms are calculated as follows:

(a) $SS_{error} = SS_{ac}$

(b) $SS_{error} = SS_{bc} + SS_{abc}$

The ANOVA table becomes:

Main treatment	A	SS_a
Blocks	C	SS_c
Error (a)		SS_{ac}
Subtreatment	B	SS_b
Interaction	AXB	SS_{ab}
Error		$SS_{bc} + SS_{abc}$
Total		SS_{total}

Split-Split Plot Design (A X B X C X D). The factors in this case are:

A = Main treatment

B = Subtreatment

C = Sub-subtreatment

D = Blocks

Consequently, there will be three separate error terms:

(a) $SS_{error} = SS_{ac}$

(b) $SS_{error} = SS_{bc} + SS_{abc}$

(c) $SS_{error} = SS_{dc} + SS_{dcb} + SS_{dca} + SS_{dcab}$

This gives the following reconstructed ANOVA table:

A	SS_a
C	SS_c
Error (a)	SS_{ac}
B	SS_b

A X B	SS_{ab}
Error (b)	$SS_{bc} + SS_{abc}$
D	SS_{d}
A X D	SS_{ad}
B X D	SS_{bd}
AXBXD	SS_{abd}
Error (c)	$SS_{dc} + SS_{dcb} + SS_{dca} + SS_{dcab}$
Total	SS_{total}

Three-Factor Randomized Blocks (A X B X C X D). Let factor C = blocks.

The error term becomes:

$$SS_{error} = SS_{ac} + SS_{bc} + SS_{dc} + SS_{abc} + SS_{acd} + SS_{bcd} + SS_{abcd}$$

Thus giving the following reconstructed ANOVA table:

A	SS_{a}
B	SS_{b}
D	SS_{d}
C (blocks)	SS_{c}
A X B	SS_{ab}
A X D	SS_{ad}
B X D	SS_{bd}
A X B X D	SS_{abd}
Error	$SS_{ac} + SS_{bc} + SS_{dc} + SS_{acd} + SS_{bcd} + SS_{abc} + SS_{abcd}$
Total	SS_{total}

2.3.1 Tests of Significance

The output of this program consists of the sums of squares and mean squares for all the main effects and interactions, together with the error mean square. In general, these main effects and interactions are tested for significance by dividing the mean square for the particular effect or interaction by the appropriate error term. The difficulty arises in the choice of the appropriate error term. A brief account of how to choose the correct error term is given below. This account is by no means comprehensive, and if the user is in any doubt as to the error term to use in his own case, he should consult H. Scheffe, The Analysis of Variance, John Wiley, 1959.

The basis for the choice of error term in ANOVA F-tests is the type of structural model used for the analysis of variance. Three models are discussed below; these should cover the majority of cases.

1. Model I (Fixed Effects). The fixed effects model is applicable when the factors used in the experiment include all possible levels for each factor, and when inferences are not made about any levels not included. Examples of this are such factors as sex, where there are only two levels possible; or training methods for teaching a specific skill; or, in a drug experiment, the treatments factor, where one is interested in the drugs used, and would not obviously want to make inferences to other drugs not included in the experiment. In the case of a fixed factor, the investigator is interested only in the levels of the variable studied in the experiment and not in any others. In this case, the computation of the F-ratio is relatively simple. The F-ratios are calculated using the error term as a divisor (for example, MS_a/MS_{error} for the A main effect; MS_{ab}/MS_{error}: AB interaction, etc.).

2. Model II (Random Effects). The random effects model applies when the experiment involves only a random sample of the set of treatments about which the experimenter wants to make inferences. For example, to study the effects of a certain drug (say alcohol) on driving skill, one would have several different levels (doses) of alcohol within the drug factor. However, all possible levels of alcohol could not be used, so one takes what is considered to be a random sample of the levels within the factor and then makes inferences about other levels.

   Another example would be the following:

   To study the effects of level of illumination on productivity in a factory, the luminance factor would be a random effects factor, since all possible levels of luminance would not be used in the experiment, but only a sample of them.

   An analysis of variance with all random effects is rarely found, and the calculation of F-ratios for this case presents some difficulties. For a two-factor model (A X B), the F-ratios are:

$$A: MS_a/MS_{ab}$$

$$B: MS_b/MS_{ab}$$

$$AB: MS_{ab}/MS_{error}$$

   For the three-factor case (A X B X C), we have the following F-ratios:

$$A: (MS_a + MS_{abc})/(MS_{ac} + MS_{ab})$$

$$B: (MS_b + MS_{abc})/(MS_{ab} + MS_{bc})$$

$$C: (MS_c + MS_{abc})/(MS_{ac} + MS_{bc})$$

AB: MS_{ab}/MS_{abc}

AC: MS_{ac}/MS_{abc}

BC: MS_{bc}/MS_{abc}

ABC: MS_{abc}/MS_{error}

In the case of the random effects model, the interactions should be tested for significance first, because if they are found to be significant, there is little point in testing the main effects for significance.

3. Model III (Mixed Model). The mixed model is probably the most common form used in analysis of variance. Here some factors are fixed, and others are random.

   The calculation of the F-ratios in this case depends on which factors are fixed and which are random. An example is given with two fixed factors and one random factor.

   A X B X C Design with Factor A a random effect:

*A: MS_{a}/MS_{error}

B: MS_{b}/MS_{ab}

C: MS_{c}/MS_{ac}

*AB: MS_{ab}/MS_{error}

*AC: MS_{ac}/MS_{error}

BC: MS_{bc}/MS_{abc}

ABC: MS_{abc}/MS_{error}

2.3.2 Job Execution

To perform an analysis of variance, the user must supply four sets of cards to the program:

1. Monitor control cards
2. Program control cards
3. Data cards
4. Table output specification cards

*Random effects

Monitor Control Cards

The monitor control cards are necessary to initiate program loading from the disk and to establish the necessary communication with the monitor. A general description of the cards may be found in IBM 1130 Disk Monitor System Reference Manual (C 26-3750).

An analysis of variance requires the following cards:

```
CC:     1  4  8       16-17
           ↓  ↓        ↓
        // XEQ ANOVA  02
        *LOCALANOVA, FMTRD, PRNTB, DATRD, STORE
        *LOCALANOV2, SDOP, MNSQ, REPRT
```

The monitor control cards do not change from job to job, but must be included with every job processed.

Program Control Cards

The program control cards communicate the data-specific parameters and output options to the program. The five card types are described below. In addition, control cards are necessary for defining the format and content of the ANOVA table (section 2.3.3).

1. Input/output units card*
2. Job-title card*
3. Option card (described below)
4. Variable format card*
5. Table generation card (section 2.3.3)

Option Card

Number of Factors (cc 1-2)

This field is punched with an integer, n, less than or equal to 4; n is the total number of factors in the experiment.

Input Mode (cc 3-4)

This field must be punched with an integer, n, which may take the values 1 or 2. If n is equal to 1, the program reads the raw data from the 1442 card reader. If n = 2, the raw data is read from the disk, where it has previously been transferred by a program using input mode number one. The data is retained until destroyed by input (mode 1) from one of the four system programs.

*See "General Operating Instructions", Chapter 1.

Transformation Switch (cc 5-6)

If the value in this field is equal to zero, the transformation program is not used. If the value is 1, the transformation program is called after each data item has been read, and before the item is stored in the appropriate storage cell. The transformation program itself is a user-written FORTRAN program, which is discussed in section 2.5.1.

Number of Levels for Factor 1 (cc 7-8)

This field must be punched with an integer, n, which indicates the number of levels in the first factor. For example, n would be equal to three for a 3X4X5 factorial design; n should be less than ten.

Number of Levels for Factor 2 (cc 9-10)

Same as cc 7-8. This field is for factor 2.

Number of Levels for Factor 3 (cc 11-12)

Same as cc 7-8. This field is for factor 3.

Number of Levels for Factor 4 (cc 13-14)

Same as cc 7-8. This field is for factor 4.

Columns 11-12 and 13-14 may be left blank for a two-factorial experiment. However, the program does not operate for fewer than two factors. All four factors, whether used or not, must be accounted for on the variable format card. The product of the levels of the factors is limited to 2000.

Analysis of Variance Option Card Summary

Column	Meaning
1-2	Number of factors
3-4	Input Mode 1 - Source data from card reader *2 - Source data from disk
5-6	Transformation Switch 0 - No transformation 1 - Transformation
7-8	Number of levels for factor 1
9-10	Number of levels for factor 2
11-12	Number of levels for factor 3
13-14	Number of levels for factor 4

*Data previously entered under mode 1 is available for mode 2 until destroyed by input (mode 1) from this or one other of the system's four main programs.

2.3.3 Analysis of Variance Table Generation

The operation of the program is designed to handle a general four-factorial design. The program will read the data, form the deviates, and accumulate the sums of squares as if all four factors were always present. As a result of this operation, certain accumulation and storage areas, in general, have cells that are not used unless all four factors are present. In forming the analysis of variance table for a particular design, the user has the option of pooling component sums of squares to form the error sums of squares specific to the design. The sums of squares are located in storage and can be accessed by the table generator cards. The table, components, and index are shown below:

Subscript	Component
1	A
2	B
3	C
4	D
5	AB
6	AC
7	AD
8	BC
9	BD
10	CD
11	ABC
12	ABD
13	ACD
14	BCD
15	ABCD

For a two-factor experiment, the sums of squares are located in cells with subscripts

1, 2, 5

For a three-factor experiment, the sums of squares are located in cells with subscripts

1, 2, 3, 5, 6, 8, 11

For a four-factor experiment, the sums of squares are located in cells with subscripts

1, 2. . . , 15

Table Generator Card Format

To print the proper component and compute the appropriate error term for a particular design, a set of table cards indicating the appropriate terms must be punched. A description of this card is given below:

Column	Meaning
1-16	Row heading for component. This field may contain any 16 or fewer characters that serve to identify the row of the table.

Column	Meaning
17-20	Print and end control. 0 or blank - Normal card. +1 - Skip to a new page and print column headings and title information before printing line for this component. -1 - Last card. No more component cards will follow. The analysis will be complete after this card is processed. The residual and total line will be printed.
21-22	Subscript of the correspondence table component, to be printed or used in accumulation of the sums of squares. For example, if this field contains a 5, the component AB will be used for either printing (the remainder of the card is blank) or accumulation.
23-24	Subscript of the cell in the correspondence table to be added to the cell used in cc 21-22. This procedure is used for adding components to form the sums of squares. For example, if cc 21-22 contained a 1 and cc 23-24 contained a 4, the printed sums of squares, mean square, and degrees of freedom would be A + AB.
25-26 27-28 . . . 49-50	The remaining two-digit fields have the same effect as cc 23-24, but are used to add additional components before printing the line. For example, if cc 21-22 contained a 2, cc 23-24 contained a 6, and cc 25-26 contained a 9, the sums of squares, mean square, and degrees of freedom would be printed as the cumulative summary of B + AC + BD.

Table Generator Card Summary

Column	Meaning
1-16	Alphameric heading for analysis of variance component.
17-20	END of table indicator 0 - More cards to follow -1 - No more cards to follow 1 - Skip to a new page before line is printed
21-22	Table component to be printed
23-24	Table component to be pooled N - Add the component with subscript N (corres. table) to the first component defined in cc 21-22.
25-26	Same as cc 23-24
27-28	Same as cc 23-24

Column	Meaning
29-30	Same as cc 23-24
31-32	Same as cc 23-24
33-34	Same as cc 23-24
35-36	Same as cc 23-24
37-38	Same as cc 23-24
39-40	Same as cc 23-24
41-42	Same as cc 23-24
43-44	Same as cc 23-24
45-46	Same as cc 23-24
47-48	Same as cc 23-24
49-50	Same as cc 23-24

2.3.4 Data Input

To set up the data for the analysis of variance, the user must identify each item of data as to its factor and level, and punch this information on a card along with the data item. Hence, each data card will have five fields, as follows:

Field	Type	Meaning
1	Integer (I)	Number of level - factor 1
2	Integer (I)	Number of level - factor 2
3	Integer (I)	Number of level - factor 3
4	Integer (I)	Number of level - factor 4
5	Floating point (F)	Observation

Disk working storage allows input of 499 observations.

The particular columns occupied by each field are arbitrary. The user describes the format of the card by means of a variable format card, which is entered into the program behind the option card. On the format card, provision must be made for all four "level" fields, even though all four fields are not necessary in the particular analysis. Figure 11 shows a sample data card from a two-factor design.

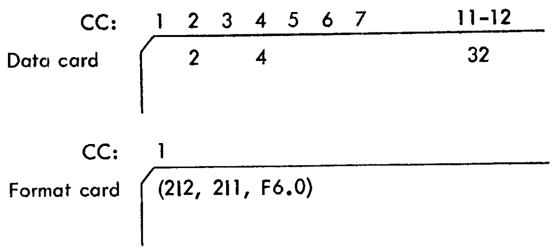


Figure 11. Sample cards from two-factorial design

In normal usage, the data items are punched one to a card, with the appropriate identification. Following the data deck, there must be an end-of-deck indicator card, which is a card containing a negative number in the first field. The order of cards is arbitrary, as the cards are rearranged in proper order before the analysis takes place.

2.3.5 Operating Instructions

A. Using the analysis of variance program when the total 1130 Statistical System has not been stored on the disk

If the user wishes to load only the set of programs that allow analyses of variance, the following programs must be compiled or assembled and stored on the disk. Each deck begins with a card punched as

//FOR

and ends with an

*STORE

card.

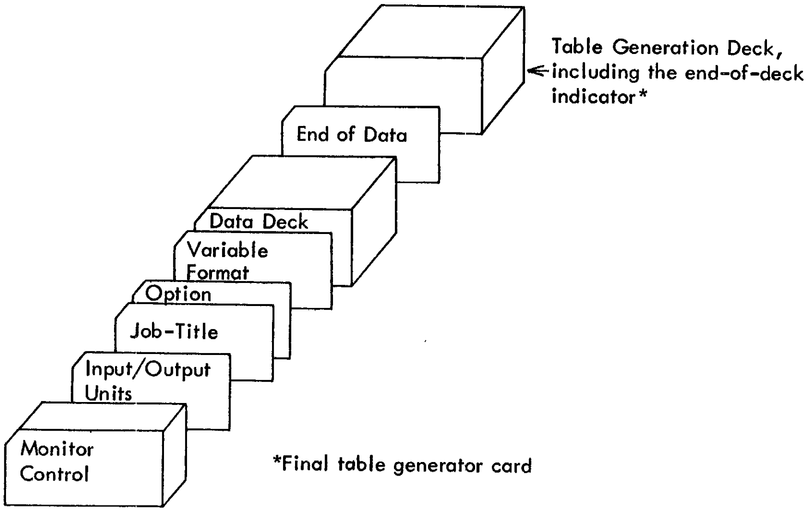
The user should use a disk containing the 1130 Disk Monitor System, as described in section 1.1. The following decks should be preceded by a cold start card, placed in the card reader hopper, and the buttons IMMEDIATE STOP (console), RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. A blank card should be placed after the last deck in the card reader hopper.

DECKS-LABELS: ANOVA-NOVA; STORE-STOR: GET-GETO; ANOV2-NOV2; SDOP-SDOP; MNSQ-MNSQ; REPRT-RPRT; *FMAT-FMAT; *FMTRD-FMRD; *DATRD-DTRD; *GMPYX-GMPY; *GDIVX-GDIV; *PRNTB-PRNB; TRAN-TRAN.

*Used in all four analysis types

B. Execution from Disk

Once the component subroutines and main calling programs are on the disk, the execution of a job requires the monitor control cards, program control cards, and data cards to be placed in the card reader. The deck should be preceded by a cold start card. To initiate processing, the buttons IMMEDIATE STOP and RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. The order in which the cards are placed in the card reader for either matrix or raw data input is shown in Figures 12 and 13.



*Final table generator card

Figure 12. ANOVA — card reader input

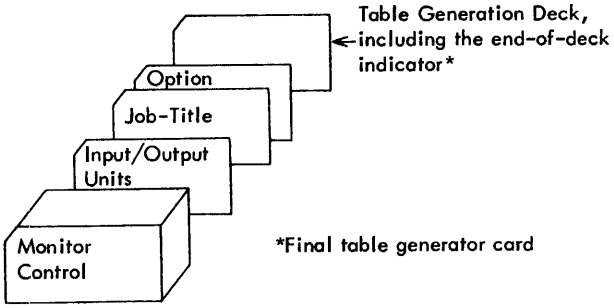


Figure 13. ANOVA — disk input

2.3.6 Sample Problem

The data for this sample problem was taken, with permission, from page 276, Statistical Theory in Research, by R.L. Anderson and T.A. Bancroft. McGraw-Hill Book Company, Inc., New York, 1952.

INPUT

```
// XEQ ANOVA    02
*LOCALANOVA,FMTRD,PRNTB,DATRD,STORE
*LOCALANOV2,SDOP,MNSQ,REPRT
020200
3333  TEST ANOVA-I
030101060302
(4I1,F4.0)
111  161
112  192
121  145
122  232
131  172
132  227
211  166
212  253
221  231
222  231
231  204
232  214
311  113
312  208
321  131
322  190
331  104
332  144
411  103
412  171
421  158
```

```
422  171
431  135
432  146
511  132
512  196
521  176
522  242
531  178
532  186
611  180
612  198
621  216
622  238
631  175
632  230
```

```
BLOCKS           0101
FERTILIZER         02
VARIETY            03
F X V              08
ERROR            -1050611

                  OUTPUT
```

```
// XEQ ANOVA   02
*LOCALANOVA,FMTRD,PRNTB,DATRD,STORE
*LOCALANOV2,SDOP,MNSQ,REPRT
```

```
          ST ANOVA-1

          NUMBER OF FACTORS                  3
          INPUT MODE                         1
          TRANSFORMATION SWITCH              1
          NUMBER OF LEVELS - FACTOR 1        6
          NUMBER OF LEVELS - FACTOR 2        3
          NUMBER OF LEVELS - FACTOR 3        2
          NUMBER OF LEVELS - FACTOR 4        0

(4I1,F4.0)
```

```
ANALYSIS OF VARIANCE TABLE FOR   6 X   3 X   2 X   0 EXPERIMENT

                             SUM OF     DEGREES OF              MEAN
   COMPONENT                 SQUARES     FREEDOM                SQUAR


BLOCKS                     24938.91         5                4987.78
FERTILIZER                  4034.00         2                2017.00

VARIETY                    17292.25         1               17292.25

F X V                       1442.66         2                 721.33
ERROR                       9896.91        25                 395.87

           TOTAL           57604.74        35

JOB COMPLETED
```

2.4 LEAST-SQUARES CURVE FITTING BY ORTHOGONAL POLYNOMIALS

Given m points (X_i, Y_i), $i=1, \ldots, m$, the (X) set not necessarily being equally spaced, this program will determine a polynomial of specified degree n or less,

$$y = a_o + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

which best approximates these points in the least-squares sense; n should not be specified greater than ten; m, of course, must be greater than n, and for practical purposes should be considerably greater than n. The program allows the number of points, m, to be as high as 149. It is difficult to envisage a requirement such that n = 7 will not suffice; however, the program has been successfully tested on polynomials of order 16 with 134 data points, with accurate results.

To maintain maximum accuracy, the program uses orthogonal polynomials, as described by G.E. Forsythe.* The process of finding the polynomial is accomplished by beginning with a first- and a second-degree polynomial and evaluating a variance criterion to determine whether the second-degree will offer a better fit than the first. If the variance criterion is satisfied within a specified tolerance, the program accepts the second-degree polynomial computed to be the best fitting polynomial. If not, the next-order polynomial is computed and compared to the second-degree. The process continues until the variance criterion is satisfied or a specified maximum degree reached. The degree of the last polynomial is assumed to be the best fitting polynomial for the data.

The essential characteristics of the method are as follows:

$$\text{Let } y = \sum_{j=0}^{n} c_j P_j \ (x) \tag{1}$$

where each P_j (x) is a polynomial of degree j.

By minimizing

$$M = \sum_{i=1}^{m} \left[y_i - \sum_{j=0}^{n} c_j P_j \ (x_i) \right]^2 \tag{2}$$

and letting

$$r_{jk} = \sum_{i=1}^{m} P_j \ (x_i) \ P_k \ (x_i) \tag{3}$$

$$S_j = \sum_{i=1}^{m} y_i P_j \ (x_i) \tag{4}$$

*Forsythe, G.E., "Generation and Use of Orthogonal Polynomials for Data Fitting with a Digital Computer", J. Soc. Indust. Appl. Math. 5, 1957; 74-78.

a set of normal equations are obtained

$$S_j = \sum_{k=0}^{n} c_k r_{jk}; \; j = 0, 1, 2, \ldots n. \tag{5}$$

Equation (5) consists of a set of n + 1 simultaneous equations with n + 1 unknowns. However, as the polynomials $P_j(x)$ are orthogonal, then,

$$r_{jk} = \begin{cases} 0 & \text{for } j \neq k \\ \sum_{i=1}^{m} P_j^2(x_i) & \text{for } j = k \end{cases} \tag{6}$$

Under this condition, the coefficients c_i can now be evaluated by

$$c_j = \frac{S_j}{r_{jj}} \quad j = 0, 1, 2, \ldots, n \tag{7}$$

The polynomials $P_j(x)$ are defined recursively by

$$\begin{aligned} P_{-1}(x) &= 0 \\ P_o(x) &= 1 \\ P_1(x) &= (x - \alpha_1)\, P_o(x) - \beta_o P_{-1}(x) \\ &\cdot \\ &\cdot \\ &\cdot \\ P_{j+1}(x) &= (x - \alpha_{j+1})\, P_j(x) - \beta_j P_{j-1}(x) \end{aligned} \tag{8}$$

where

$$\alpha_{j+1} = \frac{\sum_{i=1}^{m} x_i P_j^2 (x_i)}{\sum_{i=1}^{m} P_j^2 (x_i)} \qquad j = 0, \ldots n \tag{9}$$

$$\beta_j = \frac{\sum_{i=1}^{m} P_j^2 (x_i)}{\sum_{i=1}^{m} P_{j-1}^2 (x_i)} \qquad j = 1, 2, \ldots n - 1 \; (\beta_o = 0) \tag{10}$$

The polynomial solution vectors c_j, α_j, and β_j are then used to compute a_j, the coefficients of the equation

$$y = a_o + a_1 x + a_2 x^2 + \ldots a_n x^n \tag{11}$$

for the degree n.

In principle, the coefficients a_j could be used in equation (11) to compute the fitted values for any given argument array. However, a_j may change rapidly as n changes; therefore, it would be necessary to compute a_j with great precision. To avoid this difficulty, fitted values use c_j, α_j and β_j to compute $P_j(x)$ from equation (8), and simultaneously compute the fitted values from equation (1).

Similarly, the derivative computation uses c_j, α_j and β_j to compute the r<u>th</u> derivative of $P_j(x)$ from the recurrence relation

$$\frac{d^r P_{j+1}}{dx^r} = (x-\alpha_{j+1}) \frac{d^r P_j}{dx^r} + r \frac{d^{r-1} P_j}{dx^{r-1}} - \beta_j \frac{d^r P_{j-1}}{dx^r} \tag{12}$$

$$r = 1, 2, \ldots, n \qquad j = 0, 1, 2, \ldots n-1$$

for any <u>given</u> set of arguments.

If n and/or the range of x is large, the elements of the orthogonal vectors generated change rapidly in size; this imposes severe restrictions on accuracy. This becomes evident to the user by viewing the changes in the elements of successive vectors $P_j(X)$ as j increases, or by viewing residuals, which in this case may tend to increase rather than decrease as j increases.

To aid in circumventing this problem, the user is allowed to elect, on option, to have the program transform X to X' such that X' is in the range (-2, 2). This transformation will cause elements of $P_j(X)$ to remain approximately uniform in size as j increases. The transformation used is

$$x' = \left[4x-2\,(x_{(1)} + x_{(m)}) \right] / (x_{(m)} - x_{(1)}), \tag{13}$$

where $x_{(i)}$ is the i^{th}-order statistic from the set (x).

When this scaling is used, the following points must be considered:

1. The values of c_j and $P_j(x')$ (equation (2)) are calculated and presented using x'. Since the y's are not transformed, the y*'s (estimated y's) and residuals calculated at x' are the same as those which would have been calculated at x if c_j and P_j had been obtained without transforming.

2. The coefficients a, from

$$y = \sum_{i=0}^{n} a_i x^i$$

are actually a', from

$$y = \sum_{i=0}^{n} a_i' (x')^i \tag{14}$$

3. If, in addition to transforming x, the user elects to punch α, β, and c for later use in obtaining y*'s from a new set of x's, that is, for later use in prediction, the transformation is retained, and the new x's will be transformed as were the original x's; y*'s for these new x's will be the same as the estimated y's would have been if transformations had not been performed.

4. Derivatives are calculated using equation (14).

$$\frac{dy}{dx} = \frac{dy}{dx'} \frac{dx'}{dx}$$

5. The elements of $P_j(x')$ will now, as j increases, be of approximately the same size for all j. The elements of $P_j(x')$ are related to those of $P_j(x)$ by the factor:

$$\left[\frac{4}{x_{(m)} - x_{(1)}}\right]^j$$

2.4.1 Summary of Output

1. (x, y) for all cases (x', y if scaling is elected).
2. Predicted value of y for 3rd, 7th, ..., kth-order polynomials.
3. Residuals for 3, 7, ..., kth-order polynomials, that is, if k=13, $y-y_{pred.}$ is given for k=3, 7, 11, and 13.
4. Orthogonal polynomials of all degrees to k.
5. Polynomial solution vectors α, β, and c — used to generate orthogonal polynomials.
6. Coefficients of fitted polynomial.
7. Predicted values for externally supplied data set.
8. kth-order derivatives of the polynomial at user-specified points, where k is the order of the polynomial.
9. Scaling equation used, if required.
10. Analysis of Variance Table.

2.4.2 Job Execution

To perform a polynomial regression analysis, the user must supply three sets of cards to the program:

1. Monitor control cards
2. Program control cards
3. Data cards

Monitor Control Cards

The monitor control cards are necessary to initiate program loading from the disk and to establish the necessary communication with the monitor. A general description of the cards may be found in IBM 1130 Disk Monitor System Reference Manual (C26-3750).

The orthogonal polynomial program requires the following cards:

```
CC:  1  4   8       16
        ↓   ↓       ↓
     // XEQ POLY    02

     *LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB

     *LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
```

Monitor control cards do not change from job to job, but must be included with every job processed. The first program operated on by this system should be preceded by a cold start card.

Program Control Cards

The program control cards communicate the data-specific parameters, and output options to the program. The four possible card types are described below.

1. Input/output units card*
2. Job-title card*
3. Option cards (described below)
4. Variable format card*

TYPE I OPTION CARD

This type of option card is to be used when data is being entered into the program for the initial computation of the best fitting polynomial.

*See "General Operating Instructions", section 1.2.

Maximum Degree of Polynomial (cc 1-2)

This field should be punched with an integer, n, which is less than or equal to ten. The program attempts to fit a polynomial to the data points until the variance criterion punched in columns 17-26 is satisfied by successive-degree polynomials. If the variance criterion is not satisfied when the degree of the polynomial reaches n, the program prints a message to this effect and continues, using the solution to the nth-degree polynomial. The value of n should be less than m+1, where m is the number of the data points read.

Input Source (cc 3-4)

This field must be punched with an integer, n, which assumes a value of either one (1) or two (2).

If n is equal to 1, the data points, followed by a negative identification card, are read from the 1442 card reader. If n is equal to 2, the data points are read from the disk, having previously been transferred there by a program using input mode 1. The data on the disk is destroyed by any program using input mode 1. As noted below (Secondary Input Sources), input type 3 also destroys disk data.

Coefficients of Fitted Polynomial (cc 5-6)

In the determination of the best fitting polynomial, the computation involves only the orthogonal polynomials and the three associated vectors called the polynomial solution vectors. The orthogonal polynomials are generated from the solution vectors whenever it is necessary. Hence, the actual coefficients of the fitted polynomial are not required for evaluation of derivatives of the computation of estimated values. However, if desired, they may be printed by punching one (1) in the field. If this field is left blank or contains a zero, the coefficients are not printed.

Evaluate Derivative Switch (cc 7-8)

This field is used to indicate whether derivatives are to be evaluated at a selected list of points. If this field contains a zero or is blank, the derivatives are not computed.

If this field contains a one (1), each data point is examined to determine whether the derivative computation indicator for that point is nonzero. When a nonzero indicator is located for a particular value of x, the program evaluates the kth-order derivatives, where k is an integer punched in cc 9-10. The printout includes the 1, 2, ..., kth-order derivatives and the estimated value of y for that particular value of x.

Maximum Order Derivatives (cc 9-10)

If cc 7-8 contains a one (1), this field must be punched with an integer, k, which is less than or equal to the order of the polynomial. It indicates the maximum order derivative to be computed when a nonzero derivative computation indicator is located. The printout includes all lower-order derivatives as well as the maximum.

Predicted Values (cc 11-12)

The value of y, as estimated, is printed for each data point, x; y is also estimated and printed with the orthogonal polynomials unless the user does not elect to have them printed (cc 31-32).

Punch Solution Vectors Switch (cc 13-14)

If this field contains a one (1), the polynomial solution vectors are punched in the standard matrix format (section 2.1.4). If the field contains a zero or is left blank, the solution vectors are not punched.

To maintain maximum numerical accuracy in computing the coefficients of the fitted polynomial derivatives and estimated values, the orthogonal polynomials are used for the computation. However, to conserve storage space, the orthogonal polynomials are recomputed each time they are used. The polynomial solution vectors as functions of the data points are used as parametric vectors in this computation to avoid making more than one pass through the original data. In effect, the solution vectors are used throughout the program to represent the coefficients of the fitted polynomials. Hence, if the user expects to use the polynomial to compute additional values or derivatives, the solution vectors should be punched out. In addition to the solution vectors, the first output card includes the scaling constants required for evaluating y and derivatives at x'. If scaling is not performed, this card is still punched, and read (but not used), under input mode 3.

Variance Criterion (cc 17-26)

This field must be punched with a positive floating-point number of the form .XXXXXXXXX. The number should include the decimal point, which may be placed anywhere in the field. No blank columns are allowed.

The variance criterion is used to determine when the best fitting polynomial has been computed. The process of fitting the points involves the computation of successively higher-degree polynomials. As each degree computation is completed, a variance criterion is developed. When the difference between any two successive variances is less than the variance criterion punched in this field, the best fitting polynomial is assumed to have the degree of the last polynomial computed. If, however, this condition is not met before the maximum degree polynomial, as defined in cc 1-2, is satisfied, the maximum degree is the degree used. If the user has no feeling for the magnitude of this number, .01 may be used.

Transformation Switch (cc 27-28)

If this field is nonzero, a user-written transformation routine is called (see section 2.5.1). TRAN is called before any scaling that the user might elect to have the program perform.

Scaling Switch (cc 29-30)

If this field is nonzero, $\underline{x}$ is scaled by a linear transformation to $\underline{x}'$ such that the elements of $\underline{x}$ are in the range (-2, 2). All calculations then deal with the data set $(\underline{x}', \underline{y})$ (see section 2.4).

Polynomial and Residual Output (cc 31-32)

If this field is nonzero, only the coefficients of the polynomial $y=a_0 + a_1x + \ldots. + a_n x^n$ are presented; the residuals, y-y*, y*, and $P_j(x)$ are not listed.

This option is helpful if solution vectors are punched, for later evaluation of y at new points x*, when y at x is not desired.

Curve Fitting Type I Option Card Summary

Column	Meaning
1-2	Maximum degree of polynomial to be fitted
3-4	Input source 1 - Raw data input from card reader 2 - Raw data input from disk
5-6	Coefficients of fitted polynomial switch 0 - Do not print 1 - Print
7-8	Evaluate derivatives switch 0 - No derivatives 1 - Derivatives
9-10	Maximum order derivative
11-12	Predicted values 0 - Do not print 1 - Print
13-14	Polynomial solution vectors punch switch 0 - Do not punch 1 - Punch
15-16	Must be zero or blank
17-26	Variance criterion
27-28	Transformation switch (to user-written program) 0 - Do not call TRAN 1 - Call TRAN
29-30	Nonzero: Scale x into (-2, 2)
31-32	Nonzero: Do not print polynomials, predicted values, or residuals

TYPE II OPTION CARD

This type of option card is to be used whenever data is entered into the program with a previously computed set of polynomial solution vectors.

Degree of Polynomial (cc 1-2)

This field is used to transmit the degree of the previously computed polynomial solution vectors that are to be read by the program.

Input Type (cc 3-4)

This field must be punched with a three (3).

This program uses this number to read in the polynomial solution vectors that were punched from a previous analysis. In addition to these solution vectors, necessary scaling constants from the previous analysis are also read. It is necessary to keep all punched output from analyses in the order in which it was punched, for later input.

In effect, the solution vectors represent the coefficients of the fitted polynomial. Hence, this option is to be used when it is desired to use the fitted polynomial to compute additional estimated values and/or to compute additional derivatives for points other than those used in the initial analysis. The data points are read under the secondary input type indicated in cc 15-16. If no data points are read for evaluation of y, only coefficients are calculated. However, in this case, these have already been calculated for the previous analysis. If the secondary input type is the card reader, previously read input, which was placed on the disk, is destroyed.

Coefficients of Fitted Polynomial (cc 5-6)

In the determination of the best fitting polynomial, the computation involves only the orthogonal polynomial and the three associated vectors called the polynomial solution vectors. The orthogonal polynomials are generated from the solution vectors whenever it is necessary. Hence, the actual coefficients of the fitted polynomial are not required for evaluation of derivatives or the computation of estimated values. However, if desired, they may be printed by punching a one (1) in this field. If this field is left blank or contains a zero, the coefficients are not printed.

Evaluate Derivative Switch (cc 7-8)

This field is used to indicate whether derivatives are to be evaluated at a selected list of points. If this field contains a zero or a blank, the derivatives are not computed.

If this field contains a one (1), each data point is examined to determine whether the derivative computation indicator for that point is nonzero. When a nonzero indicator is located for a particular value of x, the program evaluates the k^{th}-order derivatives, where k is an integer punched in cc 9-10. The printout includes the 1, 2, . . . , k^{th}-order derivatives and the estimated values of y for that particular value of x.

Maximum Order Derivatives (cc 9-10)

If cc 7-8 contains a one (1), this field must be punched with an integer, k, which is less than or equal to the order of the polynomial. It indicates the maximum order derivative to be computed when a nonzero derivative computation indicator is located. The printout includes all lower-order derivatives as well as the maximum. If the order is given as zero, no derivatives are calculated.

Estimated Value Switch (cc 11-12)

This field is used to indicate whether estimated values are to be computed for the values of x read in by the programs. If this field contains a zero or is left blank, the estimated values are not computed. If a one (1) is punched, the estimated values are computed.

Secondary Input Sources (cc 15-16)

This field must be punched with an integer one (1) or two (2).

If a one (1) is entered, the data points, followed by a negative identification card, are read from the 1442 card reader, and onto the disk, destroying previously stored data. If a two (2) is entered, the data points are read from the disk. If no data points are entered, the format card is still required, and the first card succeeding the solution vector deck is read as the new data card. If this card is blank, a y at x = zero is evaluated.

If the user-written program TRAN was called, for initial analysis of the data, the user should be sure that it is called to operate on the new data.

Curve Fitting Type II Option Card Summary

Column	Meaning
1-2	Degree of polynomial
3-4	Input type 3 - Polynomial solution vectors from card reader
5-6	Coefficients of fitted polynomial 0 - No 1 - Yes
7-8	Evaluate derivatives 0 - No 1 - Yes
9-10	Maximum order derivatives

Column	Meaning
11-12	Compute estimated values 0 - No 1 - Yes
13-14	Not used
15-16	Secondary input source 1 - Raw data from card reader 2 - Raw data from disk
17-26	Not used
27-28	Transformations 0 - Do not call TRAN 1 - Call TRAN
29-32	Not used (however, x is scaled if the x's were scaled for the previous analysis)

2.4.3 Data Input

Raw data input to the program consists of a set of points (x, y) punched on cards, one set ~~poin~~t to a card, with associated identification. The general form for this input can be described in terms of individual fields for each item on the card.

Field	Type	Meaning
1	Integer (I)	Card identification. Any numeric information that serves to identify the point (x, y). The number punched in this field must be greater than zero.
2	Integer (I)	Derivative computation indicator. The program computes derivatives at any point specified in the data set. If the user wishes to have a derivative of the polynomial evaluated at the point punched on this card, this field should contain a one (1). The order of derivative to be computed is specified on the option card. If this field contains a zero (0), the derivative is not evaluated at this point.
3	Floating point (F)	The value of x. Any floating-point number is allowed. The values of x need not be equally spaced. Scaling of data may be required.
4	Floating point (F)	The value of y. Any floating-point number is allowed.

The particular card columns for each field are arbitrary as long as all four fields are present on the card.

Following the data deck, the user must include a card containing a negative integer in the identification field. This card signals the program that no more data points are to be processed.

2.4.4 Operating Instructions

A. Using the polynomial regression program when the total 1130 Statistical System has not been stored on the disk

If the user wishes to load only the set of programs that allow this type of analysis, the following programs must be compiled or assembled and stored on the disk. Each deck begins with a card punched as

//FOR

and ends with an

*STORE

card.

The user should use a disk containing the 1130 Disk Monitor System, as described in section 1.1. The following decks should be preceded by a cold start card, placed in the card reader hopper, and the buttons IMMEDIATE STOP (console), RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. A blank card should be placed after the last deck in the card reader hopper.

DECKS-LABELS: POLY-POLY; POL2-POL2; POLSQ-PLSQ;PCOEF-PCOF; PFIT-PFIT; PDER-PDER; *FMAT-FMAT; *FMTRD-FMRD; *DATRD-DTRD; *PRNTB-PRNB; *GMPYX-GMPY; *GDIVX-GDIV; TRAN-TRAN.

B. Execution from Disk

Once the component subroutines and main calling programs are on the disk, the execution of a job requires the monitor control cards, program control cards, and data cards to be placed in the card reader. The deck should be preceded by a cold start card. To initiate processing, the buttons IMMEDIATE STOP and RESET (console), START (card reader), and PROGRAM LOAD (console) should be pressed. The order in which the cards are placed in the card reader for solution vector, disk, or raw data input is shown in Figures 14, 15, and 16.

*Used in all four analysis types

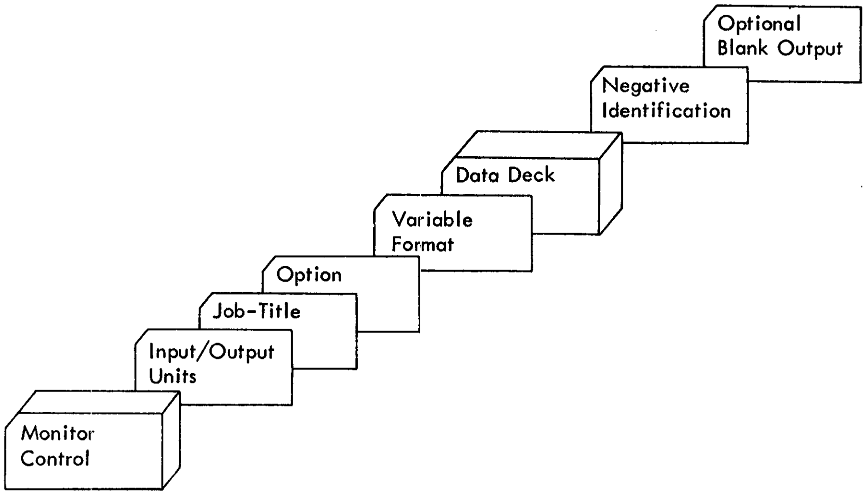


Figure 14. Orthogonal polynomial card order — card reader input

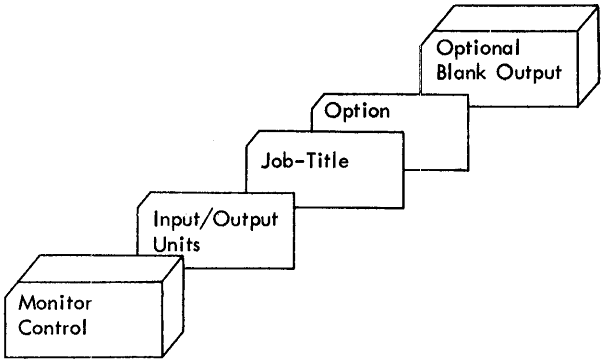


Figure 15. Orthogonal polynomial card order — disk input

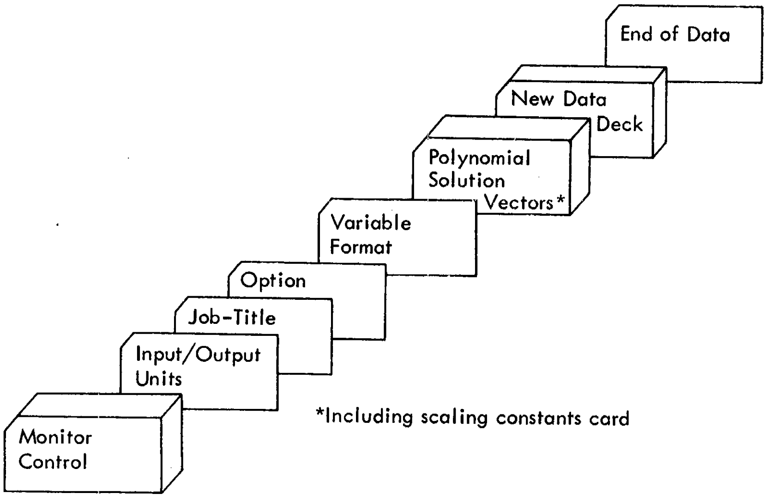


Figure 16. Orthogonal polynomial card order — solution vector input

2.4.5 Sample Problem

The data for this sample problem was taken, with permission, from page 213, Statistical Theory in Research, by R. L. Anderson and T. A. Bancroft. McGraw-Hill Book Company, Inc., New York, 1952.

INPUT

```
// XEQ POLY     02
*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
020200
1111     ORTH POLY (NO SCALING)
0201010102010100.010000000000000
(I2,I1,1X,F2.0,F3.1)
011 01011
020 02071
031 03110
040 04126
051 05147
060 06199
071 07251
080 08239
091 09231
100 10236
111 11260
120 12246
-1
```

The numbers on this first card are valid only when the user elects to scale (see section 2.4.2). When scaling is not performed, they reflect prior core status and should be ignored (that is, they can take on any value).

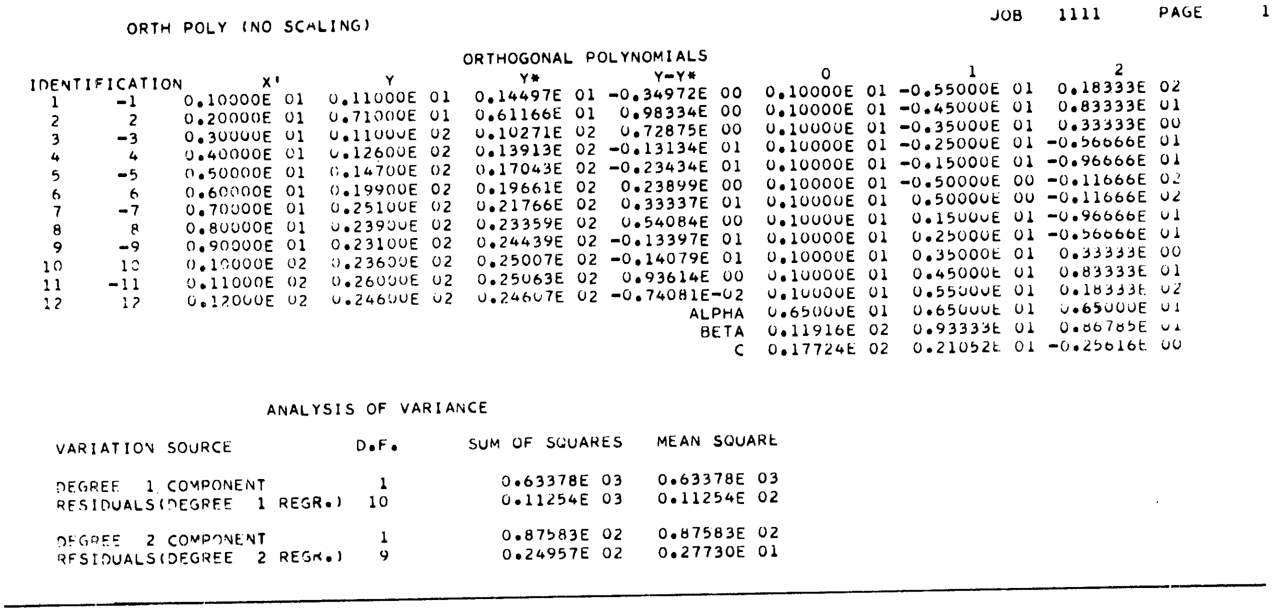
PUNCHED OUTPUT

```
 0.3636363E 00 0.1636363E 01 0
111124 1 1 0.6500000E 01 0.1191666E 02 0.1772499E 02
111124 1 2 0.6500000E 01 0.9333328E 01 0.2105245E 01
111124 1 3 0.6500000E 01 0.8678567E 01-0.2561686E 00
```

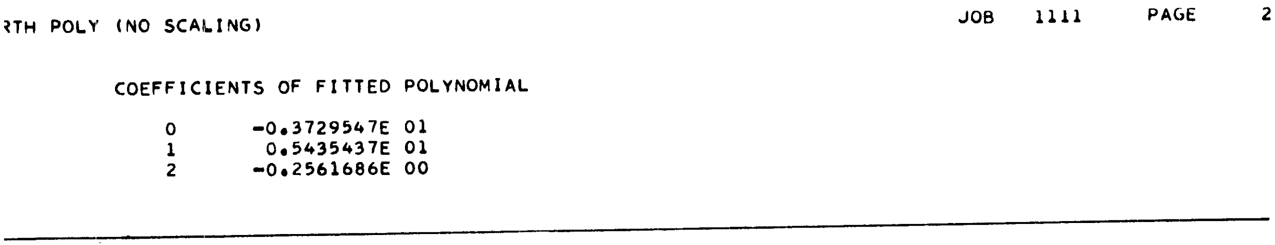
OUTPUT

```
// XEQ POLY     02
*LOCALPOL2,POLSQ,PCOFF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
```

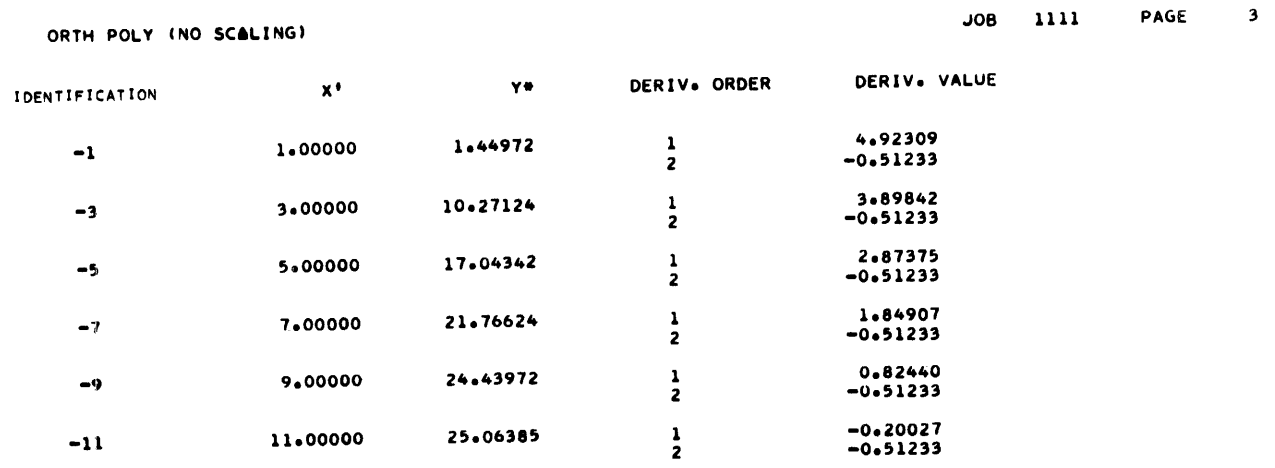
```
          ORTH POLY (NO SCALING)

           MAXIMUM DEGREE OF POLYNOMIAL        2
           INPUT TYPE                          1
           POLYNOMIAL COEFFICIENTS             1
           COMPUTE DERIVATIVES                 1
           ORDER OF DERIVATIVE                 2
           PREDICTED VALUES                    1
           PUNCH SOLUTION VECTORS              1
           SECONDARY INPUT TYPE                0

           VARIANCE CRITERION        0.010000001
           TRANSFORMATION SWITCH               0

           SCALING                             0
           IGNORE POLYNOMIAL OUTPUT            0


(I2,I1,1X,F2.0,F3.1)


X = X' (NO TRANSFORMATION)



MAX DEGREE OF POLYNOMIAL REACHED.  VARIANCE CRITERION NOT SATISFIED
```



```
                                                                               JOB    1111      PAGE      1
     ORTH POLY (NO SCALING)
                                                 ORTHOGONAL POLYNOMIALS
                                                                                  0             1              2
IDENTIFICATION          X'             Y           Y*           Y-Y*
  1       -1      0.10000E 01  0.11000E 01  0.14497E 01 -0.34972E 00   0.10000E 01 -0.55000E 01   0.18333E 02
  2        2      0.20000E 01  0.71000E 01  0.61166E 01  0.98334E 00   0.10000E 01 -0.45000E 01   0.83333E 01
  3       -3      0.30000E 01  0.11000E 02  0.10271E 02  0.72875E 00   0.10000E 01 -0.35000E 01   0.33333E 00
  4        4      0.40000E 01  0.12600E 02  0.13913E 02 -0.13134E 01   0.10000E 01 -0.25000E 01  -0.56666E 01
  5       -5      0.50000E 01  0.14700E 02  0.17043E 02 -0.23434E 01   0.10000E 01 -0.15000E 01  -0.96666E 01
  6        6      0.60000E 01  0.19900E 02  0.19661E 02  0.23899E 00   0.10000E 01 -0.50000E 00  -0.11666E 02
  7       -7      0.70000E 01  0.25100E 02  0.21766E 02  0.33337E 01   0.10000E 01  0.50000E 00  -0.11666E 02
  8        8      0.80000E 01  0.23900E 02  0.23359E 02  0.54084E 00   0.10000E 01  0.15000E 01  -0.96666E 01
  9       -9      0.90000E 01  0.23100E 02  0.24439E 02 -0.13397E 01   0.10000E 01  0.25000E 01  -0.56666E 01
 10       10      0.10000E 02  0.23600E 02  0.25007E 02 -0.14079E 01   0.10000E 01  0.35000E 01   0.33333E 00
 11      -11      0.11000E 02  0.26000E 02  0.25063E 02  0.93614E 00   0.10000E 01  0.45000E 01   0.83333E 01
 12       12      0.12000E 02  0.24600E 02  0.24607E 02 -0.74081E-02   0.10000E 01  0.55000E 01   0.18333E 02
                                                                ALPHA   0.65000E 01  0.65000E 01   0.65000E 01
                                                                 BETA   0.11916E 02  0.93333E 01   0.86785E 01
                                                                    C   0.17724E 02  0.21052E 01  -0.25616E 00


                            ANALYSIS OF VARIANCE

  VARIATION SOURCE                 D.F.         SUM OF SQUARES    MEAN SQUARE

  DEGREE  1 COMPONENT                1              0.63378E 03   0.63378E 03
  RESIDUALS(DEGREE  1 REGR.)        10              0.11254E 03   0.11254E 02

  DEGREE  2 COMPONENT                1              0.87583E 02   0.87583E 02
  RESIDUALS(DEGREE  2 REGR.)         9              0.24957E 02   0.27730E 01
```



```
ɜTH POLY (NO SCALING)                                                    JOB    1111      PAGE      2

       COEFFICIENTS OF FITTED POLYNOMIAL

           0       -0.3729547E 01
           1        0.5435437E 01
           2       -0.2561686E 00
```



```
                                                                      JOB    1111     PAGE     3
   ORTH POLY (NO SCALING)

IDENTIFICATION           X'            Y*        DERIV. ORDER     DERIV. VALUE

      -1              1.00000       1.44972          1                4.92309
                                                     2               -0.51233

      -3              3.00000      10.27124          1                3.89842
                                                     2               -0.51233

      -5              5.00000      17.04342          1                2.87375
                                                     2               -0.51233

      -7              7.00000      21.76624          1                1.84907
                                                     2               -0.51233

      -9              9.00000      24.43972          1                0.82440
                                                     2               -0.51233

     -11             11.00000      25.06385          1               -0.20027
                                                     2               -0.51233
```

```
  ORTH POLY (NO SCALING)                                                       JOB   1111      PAGE      4

IDENTIFICATION          X'              Y               Y*          Y-Y*


      1      -1     0.10000E 01   0.11000E 01   0.14497E 01 -0.34972E 00
      2       2     0.20000E 01   0.71000E 01   0.61166E 01  0.98334E 00
      3      -3     0.30000E 01   0.11000E 02   0.10271E 02  0.72875E 00
      4       4     0.40000E 01   0.12600E 02   0.13913E 02 -0.13134E 01
      5      -5     0.50000E 01   0.14700E 02   0.17043E 02 -0.23434E 01
      6       6     0.60000E 01   0.19900E 02   0.19661E 02  0.23899E 00
      7      -7     0.70000E 01   0.25100E 02   0.21766E 02  0.33337E 01
      8       8     0.80000E 01   0.23900E 02   0.23359E 02  0.54084E 00
      9      -9     0.90000E 01   0.23100E 02   0.24439E 02 -0.13397E 01
     10      10     0.10000E 02   0.23600E 02   0.25007E 02 -0.14079E 01
     11     -11     0.11000E 02   0.26000E 02   0.25063E 02  0.93614E 00
     12      12     0.12000E 02   0.24600E 02   0.24607E 02 -0.74081E-02

JOB COMPLETED
```

INPUT

```
// XEQ POLY     02
*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
020200
1111    ORTH POLY (SCALING)
0201010102010100.010000000000100
(I2,I1,1X,F2.0,F3.1)
011 01011
020 02071
031 03110
040 04126
051 05147
060 06199
071 07251
080 08239
091 09231
100 10236
111 11260
120 12246
-1
```

PUNCHED OUTPUT

```
 0.3636363E 00-0.2363636E 01 1
111124 1 1 0.3178914E-06 0.1575755E 01 0.1772499E 02
111124 1 2-0.1016575E-06 0.1234159E 01 0.5789424E 01
111124 1 3 0.3039385E-06 0.1147578E 01-0.1937265E 01
```

OUTPUT

```
// XEQ POLY     02

*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
```

```
       ORTH POLY (SCALING)                                                              JOB    1111      PAGE      0

        MAXIMUM DEGREE OF POLYNOMIAL         2
        INPUT TYPE                           1
        POLYNOMIAL COEFFICIENTS              1
        COMPUTE DERIVATIVES                  1
        ORDER OF DFRIVATIVE                  2
        PREDICTFD VALUES                     1
        PUNCH SOLUTION VECTORS               1
        SECONDARY INPUT TYPE                 0
        VARIANCE CRITERION         0.010000001
        TRANSFORMATION SWITCH                0

        SCALING                              1
        IGNORE POLYNOMIAL OUTPUT             0

(I2.I1.1X.F2.0.F3.1)


THE X VALUES HAVE BEEN TRANSFORMED TO X'=( 0.3636363E 00)*X + (-0.2363636E 01).



MAX DEGREE OF POLYNOMIAL REACHED.  VARIANCE CRITERION NOT SATISFIED
```

```
        ORTH POLY (SCALING)                                                           JOB    1111      PAGE      1

                                                     ORTHOGONAL POLYNOMIALS
IDENTIFICATION           X'              Y               Y*           Y-Y*             0             1             2
  1      -1     -0.19999E 01   0.11000E 01   0.14497E 01 -0.34974E 00       1.000000     -2.000000      2.424242
  2       2     -0.16363E 01   0.71000E 01   0.61166E 01  0.98334E 00       1.000000     -1.636363      1.101929
  3      -3     -0.12727E 01   0.11000E 02   0.10271E 02  0.72875E 00       1.000000     -1.272727      0.044078
  4       4     -0.90909E 00   0.12600E 02   0.13913E 02 -0.13134E 01       1.000000     -0.909091     -0.749309
  5      -5     -0.54545E 00   0.14700E 02   0.17043E 02 -0.23434E 01       1.000000     -0.545454     -1.278235
  6       6     -0.18181E 00   0.19900E 02   0.19660E 02  0.23902E 00       1.000000     -0.181818     -1.542697
  7      -7      0.18181E 00   0.25100E 02   0.21766E 02  0.33337E 01       1.000000      0.181817     -1.542697
  8       8      0.54545E 00   0.23900E 02   0.23359E 02  0.54086E 00       1.000000      0.545454     -1.278235
  9      -9      0.90909E 00   0.23100E 02   0.24439E 02 -0.13397E 01       1.000000      0.909090     -0.749310
 10      10      0.12727E 01   0.23600E 02   0.25007E 02 -0.14079E 01       1.000000      1.272726      0.044077
 11     -11      0.16363E 01   0.26000E 02   0.25063E 02  0.93614E 00       1.000000      1.636363      1.101928
 12      12      0.20000E 01   0.24600E 02   0.24607E 02 -0.74310E-02       1.000000      1.999999      2.424242
                                                                   ALPHA  0.31789E-06 -0.10165E-06  0.30393E-06
                                                                    BETA  0.15757E 01  0.12341E 01  0.11475E 01
                                                                       C  0.17724E 02  0.57894E 01 -0.19372E 01

READY THF PUNCH WITH BLANK CARDS AND PRESS START ON THE PUNCH AND CONSOLE.  TURN CONSOLE SWITCH 15 ON.

                         ANALYSIS OF VARIANCE

   VARIATION SOURCF                D.F.       SUM OF SQUARES    MEAN SQUARE

   DEGREE  1 COMPONFNT              1          0.63378E 03     0.63378E 03
   RESIDUALS(DEGREE  1 REGR.)      10          0.11254E 03     0.11254E 02

   DEGREE  2 COMPONENT              1          0.87584E 02     0.87584E 02
   RESIDUALS(DEGREE  2 REGR.)       9          0.24956E 02     0-27729E 01
```

```
ORTH POLY (SCALING)                                                                 JOB    1111      PAGE      2


          COEFFICIENTS OF FITTED POLYNOMIAL

             0         0.2077764E 02
             1         0.5789424E 01
             2        -0.1937265E 01
```

```
    ORTH POLY (SCALING)                                                            JOB    1111      PAGE      3

IDENTIFICATION             X'               Y*          DERIV. ORDER         DERIV. VALUE

       -1               -1.99999         1.44974             1                 13.53848
                                                             2                 -3.87453

       -3               -1.27272        10.27124             1                 10.72064
                                                             2                 -3.87453

       -5               -0.54545        17.04340             1                  7.90280

                                                             2                 -3.87453

       -7                0.18181        21.76622             1                  5.08496

                                                             2                 -3.87453

       -9                0.90909        24.43971             1                  2.26712

                                                             2                 -3.87453

      -11                1.63636        25.06386             1                 -0.55071
                                                             2                 -3.87453
```

```
        ORTH POLY (SCALING)                                                       JOB    1111      PAGE      4

     IDENTIFICATION          X'             Y             Y*          Y-Y*


           1      -1    -0.19999E 01  0.11000E 01  0.14497E 01 -0.34974E 00
           2       2    -0.16363E 01  0.71000E 01  0.61166E 01  0.98334E 00
           3      -3    -0.12727E 01  0.11000E 02  0.10271E 02  0.72875E 00
           4       4    -0.90909E 00  0.12600E 02  0.13913E 02 -0.13134E 01
           5      -5    -0.54545E 00  0.14700E 02  0.17043E 02 -0.23434E 01
           6       6    -0.18181E 00  0.19900E 02  0.19660E 02  0.23902E 00
           7      -7     0.18181E 00  0.25100E 02  0.21766E 02  0.33337E 01
           8       8     0.54545E 00  0.23900E 02  0.23359E 02  0.54086E 00
           9      -9     0.90909E 00  0.23100E 02  0.24439E 02 -0.13397E 01
          10      10     0.12727E 01  0.23600E 02  0.25007E 02 -0.14079E 01
          11     -11     0.16363E 01  0.26000E 02  0.25063E 02  0.93614E 00
          12      12     0.20000E 01  0.24600E 02  0.24607E 02 -0.74310E-02
JOB COMPLETED
```

INPUT

```
// XEQ POLY    02
*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
020200
1111    ORTH POLY (NO SCALING, SOLUTION VECTOR INPUT)
0203010102010001000000000000000
(I2,I1,1X,F2.1,F3.0)
 0.3636363E 00 0.1636363E 01 0
111124 1 1 0.6500000E 01 0.1191666E 02 0.1772499E 02
111124 1 2 0.6500000E 01 0.9333328E 01 0.2105245E 01
111124 1 3 0.6500000E 01 0.8678567E 01-0.2561686E 00
010 05000
020 55000
031 10000
-1
```

These numbers were produced in the first sample problem, and as explained there, are not used. However, the card is necessary.

OUTPUT

```
// XEQ POLY    02
*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT

*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
```

```
        ORTH POLY (NO SCALING, SOLUTION VECTOR INPUT)                    JOB    1111      PAGE      0

         MAXIMUM DEGREE OF POLYNOMIAL        2
         INPUT TYPE                          3
         POLYNOMIAL COEFFICIENTS             1
         COMPUTE DERIVATIVES                 1

         ORDER OF DERIVATIVE                 2
         PREDICTED VALUES                    1
         PUNCH SOLUTION VECTORS              0
         SECONDARY INPUT TYPE                1

         VARIANCE CRITERION        0.000000000
         TRANSFORMATION SWITCH               0

         SCALING                             0
         IGNORE POLYNOMIAL OUTPUT            0

(I2,I1,1X,F2.1,F3.0)

X = X' (NO TRANSFORMATION)
```

```
ORTH POLY (NO SCALING, SOLUTION VECTOR INPUT)                            JOB    1111      PAGE      1

          COEFFICIENTS OF FITTED POLYNOMIAL

               0        -0.3729551E 01

               1         0.5435436E 01
               2        -0.2561686E 00
```

```
    ORTH POLY (NO SCALING, SOLUTION VECTOR INPUT)                        JOB    1111      PAGE      2

IDENTIFICATION            X'              Y*          DERIV. ORDER       DERIV. VALUE

      -3                1.00000        1.44971             1               4.92309
                                                           2              -0.51233
```

```
        ORTH POLY (NO SCALING, SOLUTION VECTOR INPUT)                    JOB    1111      PAGE      3
      IDENTIFICATION         X'            Y             Y*           Y-Y*


          1       1     0.50000E 00  0.00000E 00 -0.10758E 01  0.10758E 01
          2       2     0.55000E 01  0.00000E 00  0.18416E 02 -0.18416E 02
          3      -3     0.10000E 01  0.00000E 00  0.14497E 01 -0.14497E 01

 JOB COMPLETED
```

INPUT

```
// XEQ POLY     02
*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
020200
1111     ORTH POLY (SCALING, SOLUTION VECTOR INPUT)
02030101020100010000000000000000
(I2,I1,1X,F2.1,F3.0)
 0.3636363E 00-0.2363636E 01 1
111124 1 1 0.3178914E-06 0.1575755E 01 0.1772499E 02
111124 1 2-0.1016575E-06 0.1234159E 01 0.5789424E 01
111124 1 3 0.3039385E-06 0.1147578E 01-0.1937265E 01
010 05000
020 55000
031 10000
-1
```

OUTPUT

```
// XEQ POLY     02

*LOCALPOL2,POLSQ,PCOEF,PDER,PFIT
*LOCALPOLY,TRAN,DATRD,FMTRD,PRNTB
```

```
        ORTH POLY (SCALING, SOLUTION VECTOR INPUT)                              JOB   1111      PAGE      0

         MAXIMUM DEGREE OF POLYNOMIAL      2
         INPUT TYPE                        3
         POLYNOMIAL COEFFICIENTS           1
         COMPUTE DERIVATIVES               1

         ORDER OF DERIVATIVE               2
         PREDICTED VALUES                  1
         PUNCH SOLUTION VECTORS            0
         SECONDARY INPUT TYPE              1

         VARIANCE CRITERION       0.000000000
         TRANSFORMATION SWITCH             0

         SCALING                           0
         IGNORE POLYNOMIAL OUTPUT          0


(I2,I1,1X,F2.1,F3.0)


THE X VALUES HAVF BEEN TRANSFORMED TO X'=( 0.3636363E 00)*X + (-0.2363636E 01).
```

```
ORTH POLY (SCALING, SOLUTION VECTOR INPUT)                                      JOB   1111      PAGE      1

        COEFFICIENTS OF FITTED POLYNOMIAL

           0        0.2077764E 02
           1        0.5789424E 01
           2       -0.1937265E 01
```

```
    ORTH POLY (SCALING, SOLUTION VECTOR INPUT)                                    JOB    1111      PAGE      2

IDENTIFICATION              X'           Y*         DERIV. ORDER        DERIV. VALUE

      -3                -1.99999       1.44974           1              13.53848
                                                         2              -3.87453
```

```
    ORTH POLY (SCALING, SOLUTION VECTOR INPUT)                                    JOB    1111      PAGE      3

    IDENTIFICATION          X'           Y            Y*           Y-Y*


       1      1    -0.21818E 01   0.00000E 00  -0.10758E 01   0.10758E 01
       2      2    -0.36363E 00   0.00000E 00   0.18416E 02  -0.18416E 02
       3     -3    -0.19999E 01   0.00000E 00   0.14497E 01  -0.14497E 01

JOB COMPLETED
```

2.5 GENERAL NOTES ON THE PROGRAMS

2.5.1 Transformations

This feature was added to aid users who are familiar with programming (see the manuals 1130 FORTRAN Language (C26-5933) and 1130 Disk Monitor System (C26-3750)) in adding transformation capability to the system. Currently, a subroutine TRAN is included in the package, and is called on option by each main program, subsequent to the reading of each observation. The current routine returns to its calling program immediately.

In implementing such a subroutine, the following points should be considered:

1. In the regression and factor analysis programs, the observation (row) X is in COMMON storage, and can be reached by use of the COMMON statement in the user-written program. The row X contains one observation on $X_1, \ldots, X_K$, and TRAN could be written using the row X as an argument.

2. In the orthogonal polynomial program, TRAN is called after each reading of x_i, y_i, which are elements of vectors X, Y, in COMMON. TRAN could have arguments x_i and/or y_i, or could use the COMMON statement.

3. For the analysis of variance program, TRAN should include the argument DATA, containing the observation.

4. If a large transformation program is prepared by the user, storage requirements may call for the use of LOCAL monitor facilities.

5. Transformations that modify the number of variables in the observation require modification to the program supplying and analyzing the data. Such modifications require programming knowledge of the package. For example, if one originally entered ten variables, and wished to transform the sum of four of them into one column of the observation matrix, the sum should be placed in the column of one of the original variables, and the program would have access to the resulting ten variables for its analysis. In this specific instance, the program would exit because of a singularity.

2.5.2 Notes on Correlation and Eigen Analysis

The regression and factor analysis programs contain options that, in proper combination, cause program termination when the correlation matrix and the latent roots and vectors have been calculated. For example, the "no print" option (cc 25-26 of the option card) used with the option for printing the correlation matrix gives this facility in the regression program.

With the matrix input option to factor analysis, and using option 2 for the number of factors, rotation option 0, and communality option 0, eigenvalues of matrices can be obtained. However, the number of eigenvectors is limited to ten, the maximum number of rotatable factors.

2.5.3 Punched Matrix Output

Matrix Number	Dimension	Name
1	N x N	Raw cross products matrix
2	N x N	Adjusted cross products matrix
3	N x N	Variance-covariance matrix
4	N x N	Correlation coefficients matrix
5	N x N	Characteristic vectors
6	N x K	Principal axis factor matrix
7	K x K	Orthogonal transformation matrix
8	N x K	Orthogonal factor matrix
9	K x K	Transformation to oblique reference structure matrix
10	N x K	Oblique reference vector structure matrix
11	K x K	Correlations among oblique reference vectors
12	N x K	Oblique reference vector pattern matrix
13	K x K	Correlations between reference vectors and primary factors
14	N x K	Oblique primary factor structure matrix
15	K x K	Correlations among oblique primary factors
16	N x K	Oblique primary factor pattern matrix
17	N x K	Factor score regression coefficients
21	1 x 1	Number of cases

The following matrices include two or three vectors, which are punched as column vectors where the column dimension indicates the number of elements in the vector.

Matrix Number	Dimension	No. of Elements on Each Card	Meaning
22	N x 2	2	Raw sums, raw sums of squares
23	N x 2	2	Means, standard deviations
24	*N x 3	3	Alpha, Beta, C (orthogonal polynomial solution vectors)

In the above: N = number of variables
K = number of rotated factors
*N = order of the polynomial

2.5.4 Scaling

The programs in the 1130 Statistical System allow large data sets and a general input format. Thus, there is a possibility that scaling will be necessary.

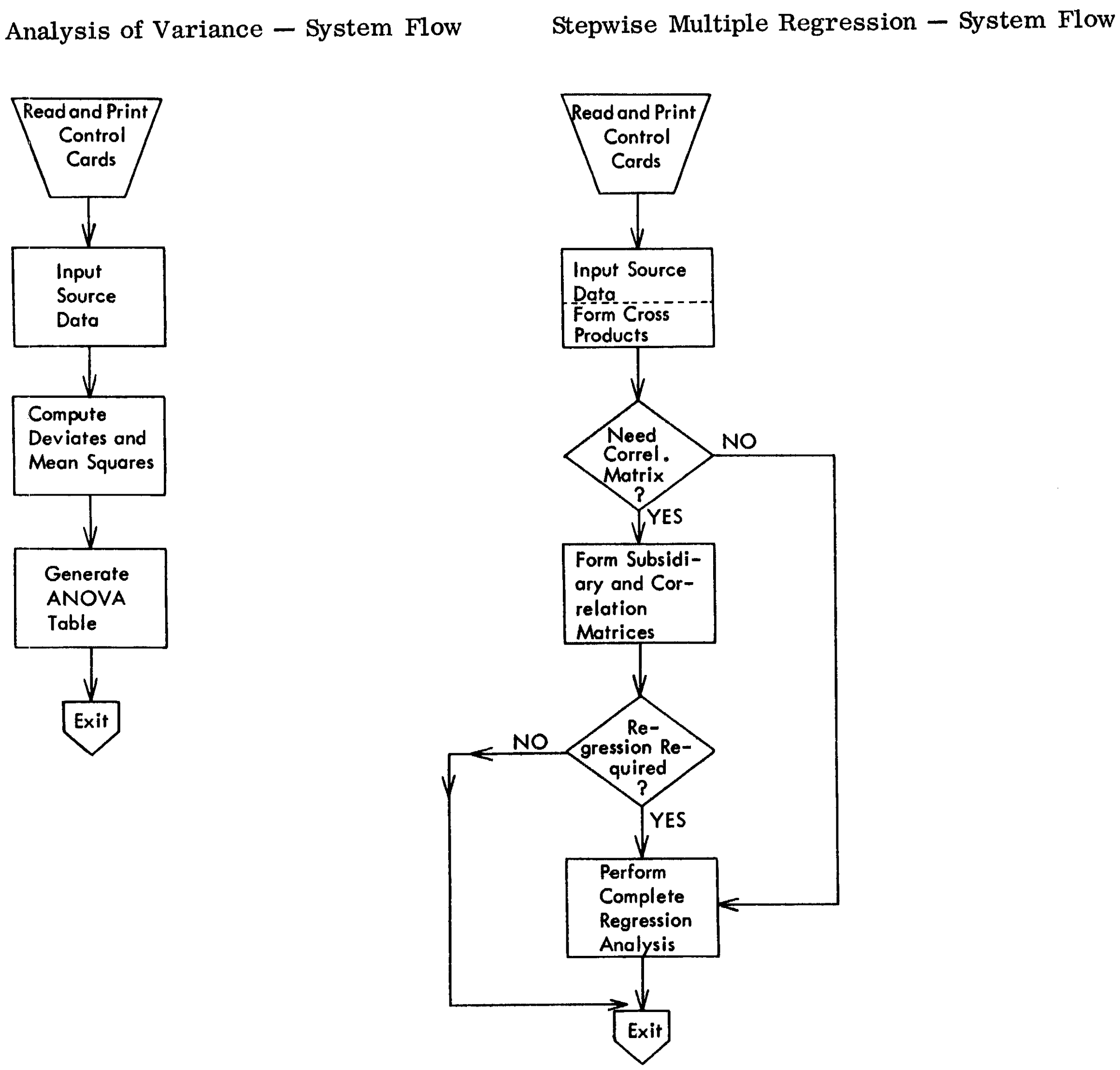
In regression and factor analysis, a pooling option is allowed that uses the raw cross products matrix. If the number of observations is large, or if some observed variable readings are quite large, some inaccuracies may become evident in this matrix. Sometimes, scaling by use of the Format statement can aid in the solution of this problem. In other situations, a transformation of the variable may help. It is also possible that scaling should take place before data entry.

In orthogonal polynomials, if the order of the polynomial is high, and/or the range of x is large, the elements of the polynomials will change rapidly in magnitude (as the order of the polynomial increases) so as to even exceed the range of the floating-point number, resulting in underflow or overflow. If data entered into this program is such that this happens, as evidenced, for example, by the residuals, the user can elect to transform or scale the dependent and/or independent variables. A useful option in this program automatically scales the independent variable into a range such that the magnitudes of the successive polynomial elements are approximately uniform.

In summary, the programs in this system are data-dependent, as is the case for many computer programs. In some programs, definite accuracy characteristics can be stated. In data-dependent routines, these statements are difficult to make.

CHAPTER 3: GENERAL FLOWCHARTS

The following charts (Figures 17, 18, 19) describe, generally, the programs in this package. More detailed flowcharts, and listings, are available in the Systems Manual for this package. The Systems Manual is not distributed with this program unless specifically requested.


Analysis of Variance — System Flow

Stepwise Multiple Regression — System Flow

Read and Print Control Cards
Input Source Data
Compute Deviates and Mean Squares
Generate ANOVA Table
Exit
Read and Print Control Cards
Input Source Data
Form Cross Products
Need Correl. Matrix ?
NO
YES
Form Subsidiary and Correlation Matrices
Regression Required ?
NO
YES
Perform Complete Regression Analysis
Exit

Figure 17. Analysis of variance and stepwise multiple regression

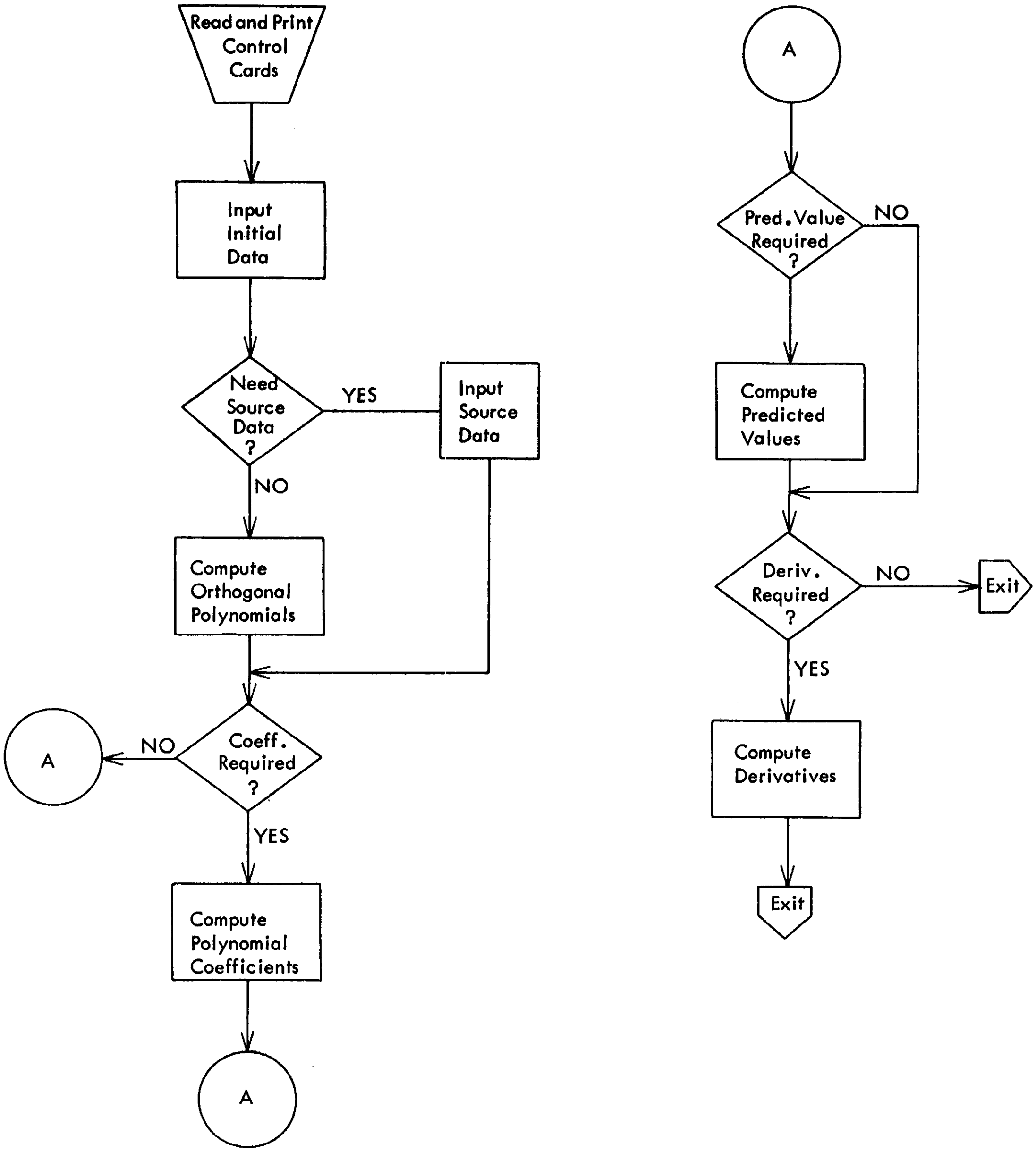


Figure 18. Curve fitting with orthogonal polynomials

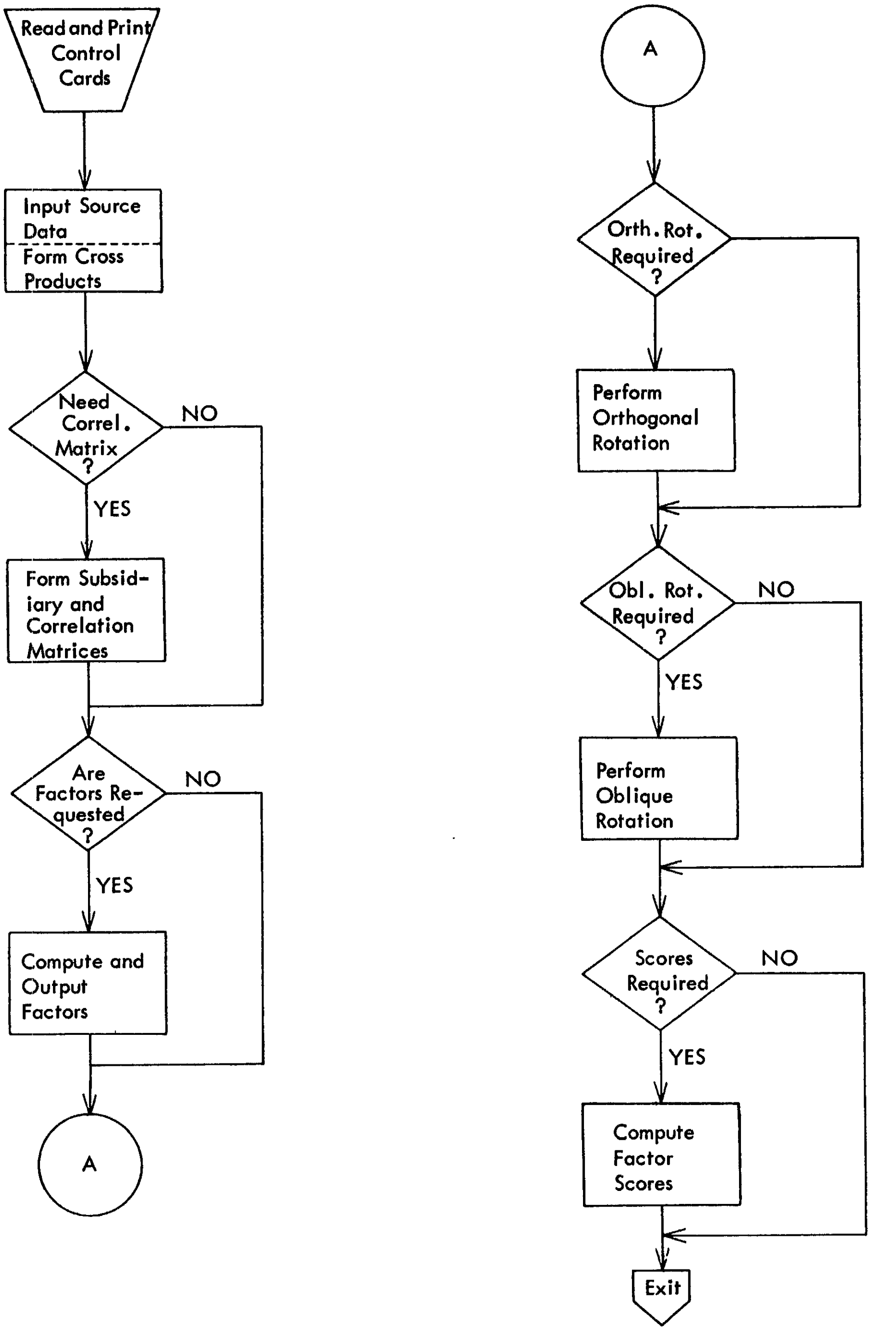


Figure 19. Factor analysis

CHAPTER 4: SAMPLE PROBLEM TIMING

The table below gives times for each sample problem, from the reading of the first monitor card to the end of the output listing.

The 1132 Printer was used as the output device.

Problem	Time (min:sec)
Regression analysis (card input)	6:02
Regression analysis (correlation matrix input)	3:04
Orthogonal polynomials (card input, no scaling)	2:35
Orthogonal polynomials (cards, scaling)	2:35
Orthogonal polynomials (solution vector input, no scaling)	2:00
Orthogonal polynomials (solution vector input, scaling)	2:00
Analysis of variance	2:14
Principal components analysis (card input)	5:48
Factor analysis (correlation matrix input)	3:45

CHAPTER 5: ERROR MESSAGES

Following is a list of error messages presented to the user on the (optional) printer or the typewriter:

Common to All Programs

1. AN ILLEGAL CHARACTER HAS BEEN ENCOUNTERED IN COLUMN (N) OF THE ABOVE FORMAT CARD. CHANGE CARD AND RERUN JOB.

   Action: Correct the format card and rerun the job. See section 1.2 (Variable Format Card).

2. AN ILLEGAL CHARACTER HAS BEEN ENCOUNTERED IN APPROXIMATELY COLUMN (N) OF THE ABOVE DATA CARD. CHANGE CARD AND RERUN JOB.

   Action: The format card and/or the data card is in error. Correct the card(s) and rerun the job. See sections 1.3(1) and 1.2 (Variable Format Card), and the data input section pertaining to the particular analysis being run.

3. INVALID INPUT OPTION. JOB TERMINATED.

   Action: Data input mode is not 01, 02, or 03. See the section discussing the option card (input mode) for the particular analysis being run.

Common to Regression and Factor Analysis

4. CARD (ID) IS OUT OF SEQUENCE. RERUN JOB.

   Action: Check sequence number of card, revise, and rerun job. See section 2.1.3 or 2.2.5.

Regression Messages

5. MEAN SQUARE NONPOSITIVE. JOB TERMINATED.

   Action: A format specification error could have caused data to be converted incorrectly, or an ill-conditioned matrix (for example, one with high correlations between independent variables) could have caused inaccuracy in the inversion of the correlation matrix or in the calculation of mean squares.

6. NO MORE DEGREES OF FREEDOM. JOB TERMINATED.

   Action: The number of parameters being estimated is larger than the number of observations. Increase the number of observations, or accept a model with fewer parameters.

7. NO MORE VARIABLES SATISFY THE VARIANCE CRITERION. JOB TERMINATED.

   Action: Modify the variance criterion (section 2.1.2), or accept one of the models produced.

IBM

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, N.Y. 10601
(USA Only)

IBM World Trade Corporation
821 United Nations Plaza, New York, New York 10017
(International)