Technical Memorandum
HuF-TM-3

August 1, 1972

ANALYSIS OF VARIANCE PROGRAM (ANOVA)

Jerold S. Lower

Highway Safety Research Institute
University of Michigan
Ann Arbor

TABLE OF CONTENTS

I. INTRODUCTION

This program was originally adapted by William L. Carlson, Systems Analysis Group, HSRI, from material contained in Hartley (1962). Jerold S. Lower, Human Factors Group, HSRI made extensive revisions to the program, including variable formatting of input data, transformations, automatic selection of error terms, provision for 3- and 4-way tables and printout of all cell means, and repeated analyses without reloading, as well as optimizing the execution time, reformatting of the printed output, and generation of this write-up. Carole D. Hafner, Computer Services Group, HSRI, wrote the routines for performing Newman-Keuls and Tukey tests on main-effect and interaction means.

II. GENERAL DESCRIPTION

ANOVA is a general, flexible program to perform analysis of variance on a wide variety of experimental designs. The program can be used with equal cell frequencies, or to perform unweighted-means analyses using unequal cell frequencies. (For the general method of performing unweighted-means analyses and the circumstances in which they are valid, see Winer [1962, pp. 222-224].) It permits a large number of optional features to be selected by the user. It accepts many types of crossed and nested designs, and selects the correct error term automatically, in accordance with principles described later in this write-up, to form the necessary $\underline{F}$ ratios. Significance tests may be performed on _post hoc_ comparisons between main-effect means, either on an overall basis, within each level of a second variable, or within each combination of levels of two other variables. Either the Newman-Keuls method or the Tukey (b) method (Winer, 1962, pp. 79-92) may be selected. A complete table of cell means and tables of means for any desired two-, three-, or four-way interactions may be printed out at the option of the user. (Main-effect means are always printed.) Data may be transformed according to several mathematical functions. Multiple analyses may be performed on the same (or different) data with a single $RUN command, to save program loading costs.

III. RESTRICTIONS

1. The number of independent variables may not exceed ten.

2. The number of cells in the design may not exceed 10,000. There is no limitation of the number of observations per cell.

3. The number of levels of each independent variable must be such that, if we let k_i equal $1+n_i$, where n_i is the number of levels of variable_i, the product of all the k_i is less than or equal to 60,000. No variable may have more than 99 levels.

4. Experimental designs must be balanced and have no missing cells.

5. Only one independent variable may be a random factor.

6. Only one variable may be a nested factor. This factor must be random and may be nested under not more than two fixed factors. The number of levels of the nested factor must be constant for all factors or combinations of factors under which it is nested.

While this program is designed to accommodate a large majority of the types of balanced designs found in the behavioral sciences, it can also be used for other designs if they are capable of being rationalized into complete factorial designs (see, for example, Clifford, 1968; Bogartz, 1968). For instance, designs with one factor nested under three or more others can be analyzed by this method, though at some additional cost due to the manual computation required to construct the resulting variance table.

IV. RUNNING THE PROGRAM

This program is currently stored under user number SAYM, in the read-only file ANOVA*. It is designed for either batch or

---

*It is possible that at some future time it will become necessary to use a different user number. In this case, the new number may be obtained by calling the HSRI Human Factors Group, 764-4158.

terminal operation. In batch mode, the set-up to be described below may be placed on cards and submitted along with the data deck, or may be stored in a file prior to the run. In terminal mode, each line of the set-up is prompted for by the program. Data is input via logical unit 2, and set-up via logical unit 4. Results are printed out on logical unit 3, and prompt lines to the terminal are issued via logical unit 5. If the set-up is submitted on cards along with the data for batch use, the required run command is:

$RUN SAYM:ANOVA 2=*SOURCE*, 3=*SINK*, 4=*SOURCE*, 5=*DUMMY*

If the set-up has been previously placed in a file the command should be:

$RUN SAYM:ANOVA 2=*SOURCE*, 3=*SINK*, 4=SETUPFILE, 5=*DUMMY*

and if the data has also been previously placed in a file, the name of this file may be substituted for *SOURCE* in the specification for logical unit 2. For terminal use the command should be as follows:

$RUN SAYM:ANOVA 2=DATAFILE, 3=BULKFILE, 4=*SOURCE*, 5=*SINK*

(Simple analyses having a small volume of output may be written out directly on the terminal by assigning 3=*SINK*.) In either terminal or batch mode, if multiple analyses are to be performed on the same set of data within one run, the data must be in a file and this file may not also contain the set-up.

The set-up file or deck should consist of the following cards (or lines): (These will be prompted for in terminal mode.)

1. Label. A heading consisting of 80 characters which will appear at the top of the printed output.

2. Cell Design. This should contain the number of independent variables, in columns 1-2, and the number of replications per cell, in columns 3-4. For unequal cell frequencies, this parameter should be entered as -1. The next two columns should contain the number of levels of the first independent variable. The next column (7) is blank, followed by a 4-character name or

label for this variable in columns 8-11. The next two columns (12-13) contain the number of levels of the second independent variable. The next column (14) is blank, the next four columns (15-18) contain the name or label for the second independent variable, etc.* In some rare instances in which the raw data values are very large or very small, the magnitude of the results printed out may be inappropriate for the formats used. In these cases, a scale factor or multiplier may be entered in columns 75-79 of this card. The decimal point, unless explicitly entered, is assumed to be at the end of this field. All input data will then be multiplied by this value before analysis. If this field is blank (the normal case) no such conversion will occur.

If a multiplier is used, then the output data will need to be reconverted if they are to be expressed in the original units. Means should be divided by the multiplier, while sums of squares and mean squares should be divided by the square of the multiplier. Degrees of freedom and *F*'s, of course, remain unchanged. If both a multiplier and a transformation are elected, the multiplication takes place before the transformation.

3. Independence of Replications. If within cell replications represent independent observations, enter "1" in column 2; if they represent correlated observation (e.g. repeated measures on the same subjects), enter "2." (Omit if there is only one score per cell.)

4. Random Factor Design. If the design involves only fixed factors, enter "2" in column 2; if one factor is random, enter "1."

---

*In general, for designs in which each subject receives each combination of one or more treatment factors (nested designs and others which employ a "subject pseudofactor") subjects must be explicitly entered into the design card as an independent variable. In completely crossed designs the number of levels equals the number of subjects; in nested designs it is the number of subjects per group.

5. Identification of Random and Nesting Factors. (Omit if the design involves only fixed factors.) The letter denoting the random factor should be entered in column 1. If this factor is nested under other factors the letter for the first such factor should be entered in column 3 and for the second such factor in column 5.

6. Replication Code. If the data deck is sorted in normal fashion (see section V.) with all replications for a given cell occurring consecutively, a "0" should be entered in column 1. If the optional sort is employed, with all cell scores for the first replication occurring consecutively, then all for the second replication occurring consecutively, etc., a "1" should be entered in column 1. (Omit if cell frequencies are unequal or 1.)

7. Number of Observations Per Data Group. The data group referred to here ordinarily will consist of one card, but in special circumstances, such as the case in which data for the several replications within one cell occupy more than one card, this data group may contain as many cards (or card images) as necessary, subject to the restrictions that the maximum number of observations in one such group is 80. Enter right justified in columns 1-5.

8. Format of Data Group. This card should contain a standard Fortran F-type format declaration describing the format of the input data within the data group referred to previously. This format will be used in reading in data. Parentheses around the declaration must be supplied, and those portions not enclosed in parentheses (statement label and word FORMAT) are to be omitted.

9. Transformation Code. The data to be analyzed may be subjected to several types of transformations. The data are transformed at read-in time, and all output from the program is in terms of the transformed, rather than the original, scale of measurement. The following transformations are provided.

Code Number	
0	No Transformation
1	Square Root
2	Natural Logarithm
3	Reciprocal
4	Arc Sine

(Enter the appropriate code in column 1)

In each case, the specified function is evaluated with respect to the number contained in the input data and the result is entered in storage to be used in the analysis of variance.

If the arc sine transformation is selected, an additional line of information will be prompted for at this point. This consists of two items. The number of discrete, independent events represented by each data point should be entered in columns 1 to 5. A code in column 10 should be set to 0 if the raw data are proportions, and 1 if the raw data are frequencies. For further information on this transformation see Winer (1962, p. 221). If the data are frequencies, the program computes the proportion automatically. The second of the two formulae given by Winer is used.

10. Augmenting Factor. In the case of several of the above transformations, such as the reciprocal and logarithmic transformations, it is often desired to add a small constant to each score before transformations, to avoid zero values in the data to be transformed. If such augmentation is desired, this constant should be entered in column 1 to 5. The decimal point is assumed to be after column 5 unless it is explicitly entered. If this augmentation is not desired, this card should contain blanks or zeros in columns 1 to 5. (Omit if no transformation.)

If the set-up is on cards along with the data, the data should follow here. The set-up then resumes as follows:

11. Cell Mean Card. Should contain "01" in columns 1-2 if a complete table of cell means is desired; "02" otherwise.

12. Post Hoc Comparison Cards. For each such set of comparisons desired a card with a "1" in the second column should be inserted here followed by a card containing the letters for the variables on which comparisons are to be performed. On this second card, the letter for the variable on which comparisons are to be made should appear in column 1. If these comparisons are to be made within levels of another variable or within combinations of levels of two other variables, the letter for the first such variable should appear in column 3 and the letter for the second such variable, if used, should appear in column 5. A third card specifying the type of test must follow. For Newman-Keuls tests it should contain a "1" in column 2; for Tukey (b) it should contain a "2." This procedure should be repeated for each such comparison wanted. When no further comparisons are wanted, the series should be terminated with a card having "2" in column 2. Note: A variable having more than 15 levels may not be used in post hoc comparisons.

13. Table Cards. If a 2-, 3-, or 4-way table of means is desired for the combinations of 2 to 4 independent variables, a card should be inserted here with a "1" in column 2. Note that if a given set of variables has been used in performing post hoc comparisons, a table for that set has automatically been produced. The variable whose different levels will form the columns of the table should be entered in column 1 of the next card. The variable whose levels will form the rows of the table should be entered in column 3. The variable (if any) whose levels will form the minor blocks of the table should be entered in column 5, and the variable whose levels will form the major blocks of the table should be entered in column 7. Letters should be used. For greatest speed and compactness of printing, the

factor with the greatest number of levels among those involved in the table should normally be used as the column factor. If this factor has more than eight levels, it will automatically be interchanged with the first factor in the list which has eight or less. If no such exchange is possible each block of the table will be broken into eight-column sub-blocks. Except for the above limitations, table layout is at the discretion of the user and some other layout may result in a table easier to interpret in a given situation.*

This procedure, like the post hoc comparisons procedure, should be repeated until no further tables are wanted. The series should then be terminated with a card having "2" in column 2. This terminates the analysis.

14. Repeat Analysis. If another analysis is desired, enter "1" in column 2; otherwise enter "2" to terminate the job.

15. Rewind Data Set. Entering "1" in column 2 causes the data set specified as logical I.O. unit #2 in the $RUN command to be rewound. If this data set is a disk file, this will cause the succeeding analysis to be done on the same data as used in the initial analysis (a new variable may be selected, however, by altering the input format). If this data set is a tape, it will be rewound to the beginning of the tape. Thus this option should not be used when the data are on tape unless it is desired to use the first file on the tape. It can not, of course, be used when the data are on cards. If the subsequent analysis is to be run on new data which follow the data for the preceding analysis in the deck or file, enter "2."

16. Label. Enter a new heading for the subsequent analysis in columns 1-80.

---

*To print tables of more than 5 columns on a Western Electric model 33 or 35 teletype or similar terminal will require setting %LEN=132. Needless to say, the resultant "wrapping around" of output lines will result in substandard appearance and readability.

17. Same Design. If the design to be used in the subsequent analysis is the same as that in the preceding one (same number of factors and replications, same name and number of levels for each factor), enter "1" in column 2 and then return to item 3 of this list to set up the new analysis. If a different cell design is desired, enter "2" in column 2 and return to item 2 of this list to set up the new analysis.

V. SORTING OF DATA

1. Equal Cell Frequencies. The data must be arranged within the deck or file in correct order. Two options are available. The two options are shown schematically in the diagrams below. In these diagrams the letter

Standard Sort (Replication Code = 0)

1.	$A_1B_1C_1R_1$	A_1	B_1	C_1
2.	$A_1B_1C_1R_2$			
3.	$A_2B_1C_1R_1$	A_2		
4.	$A_2B_1C_1R_2$			
5.	$A_1B_2C_1R_1$	A_1	B_2	
6.	$A_1B_2C_1R_2$			
7.	$A_2B_2C_1R_1$	A_2		
8.	$A_2B_2C_1R_2$			
9.	$A_1B_1C_2R_1$	A_1	B_1	C_2
10.	$A_1B_1C_2R_2$			
11.	$A_2B_1C_2R_1$	A_2		
12.	$A_2B_1C_2R_2$			
13.	$A_1B_2C_2R_1$	A_1	B_2	
14.	$A_1B_2C_2R_2$			
15.	$A_2B_2C_2R_1$	A_2		
16.	$A_2B_2C_2R_2$			

"R" refers to replications, whereas the letter "A" refers to levels of the first factor, "B" refers to levels of the second factor, etc. In the first (and usual) option, all replications within a given cell appear consecutively within the input data. The cells are arranged as follows: holding all factors except the first one constant at their first level, the cells for the various levels of the first factor appear in natural order. The second factor is then advanced to its second level, and the levels of the first factor are again repeated in order. This is followed by the same procedure at the third level of the second factor, etc., until all levels have been exhausted. The third factor is then advanced to the second level, etc. Note that the observations which appear consecutively in the input data need not necessarily occupy consecutive positions on their respective cards, nor be on consecutive cards, provided that the pattern of the data format is a regularly recurring one. By use of appropriate format statements it is possible to skip over unwanted data.

In the second option, which may be desirable in certain special circumstances (e.g. when the replications represent repeated measures taken with substantial time periods intervening), the cells of the design are arranged in the same order as in the first option, but the cell data from the first replication occur in a block, then the entire sequence of cells is repeated to form a similar block for each subsequent replication. The examples shown assume a 2 X 2 X 2 design with 2 replications. The letters A, B, and C represent the independent variables, and R represents replications, while the numbers represent levels of the factors.

Optional Sort (Replication Code = 1)

No.	Cell	B	C	R
1.	$A_1B_1C_1R_1$	B_1	C_1	R_1
2.	$A_2B_1C_1R_1$			
3.	$A_1B_2C_1R_1$	B_2		
4.	$A_2B_2C_1R_1$			
5.	$A_1B_1C_2R_1$	B_1	C_2	
6.	$A_2B_1C_2R_1$			
7.	$A_1B_2C_2R_1$	B_2		
8.	$A_2B_2C_2R_1$			
9.	$A_1B_1C_1R_2$	B_1	C_1	R_2
10.	$A_2B_1C_1R_2$			
11.	$A_1B_2C_1R_2$	B_2		
12.	$A_2B_2C_1R_2$			
13.	$A_1B_1C_2R_2$	B_1	C_2	
14.	$A_2B_1C_2R_2$			
15.	$A_1B_2C_2R_2$	B_2		
16.	$A_2B_2C_2R_2$			

2. Unequal Cell Frequencies. If cell frequencies are unequal, the data must be sorted in the standard fashion. (All scores for a given cell grouped together.) The data deck or file must contain a sufficient number of fields for each cell, including the last cell, to accommodate the largest number of replications found in any one cell. Within each cell the fields should be punched consecutively until all scores have been entered, and those remaining should be left blank or filled with zeros.

The data must be preceded by the following:

a. A card containing the maximum number of scores in any cell, right justified in columns 1-4.

b. A series of cards containing the number of scores in each cell, arranged in the same order as the cells are arranged, twenty per card, right-justified in four column fields.

VI. SELECTION OF ERROR TERM

The program has been designed to automatically construct a variance table and compute *F*-ratios for each main effect or interaction which is testable. The methods of selecting the error term(s) are derived from Winer (1962) and McNemar (1962). They are summarized below.

1. *Designs Involving Fixed Factors Only*.

A. *Designs Without Replication*. The highest-order interaction is used as the error term. Since this contains the desired error term confounded with any real *n*-way interaction that may exist among the *n* factors, it will lead to too few significant F's unless this interaction is zero. (The significance of the interaction is not testable.) A warning about this confounding is printed out. Of course, if any of the *F*'s are significant, one may assert with more than usual confidence that the effect(s) represented are not due to chance; however the interpretation of a negative result is more tenuous than usual.

B. *Designs with Replication*. If replications are independent, the "within cells" means square is used as the error term for all other mean squares.

If replications are correlated, the design is treated in the same fashion as a design without replication except that an *F* is calculated for the highest-order interaction, using the "within cells" mean square as denominator. This *F* is meaningless as a measure of the significance of this interaction, since it is ordinarily confounded with individual differences. In the usual situation, the denominator of this *F* is variance attributable to unreliability of measurement, and since real individual differences usually exist, this *F* may ordinarily be expected to be significant. A small value for *F* here should be taken to indicate unreliability of measurement and thus cast doubt on the precision of the experiment. The other *F*'s are interpreted in the same fashion as for designs without replication.

2. Designs Involving a Random Factor.

A. Crossed Designs. The random factor is not testable for significance, nor are any of the interactions between the random factor and fixed factors except the (*n*-1)-way interactions, where *n* is the total number of factors. The denominator for *F* in these cases is the *n*-way interaction. This is also used as numerator to form an *F* with the "within cells" mean square as denominator when there is within-cell replication. This latter *F*, like the one discussed in section VI.1.B., is interpretable only in terms of a comparison between variance due to individual differences and that due to experimental error.

Each main effect and interaction involving only fixed factors is tested for significance, using the interaction between this factor or combination of factors and the random factor as error term.

B. Nested Designs. In this discussion a factor nested under one or more other factors will be referred to as a nested factor, while the factors under which it is nested will be referred to as nesting factors. All other factors will be referred to as crossed factors.

The nesting factors and their interaction (if any) will be tested for significance against pooled error term consisting of the variances due to the nested factor and to its interactions with the nesting factors and their combination (if any). Thus if the nesting factors are A and B and if the nested factor is C, the pooled error term for testing A, B, and AB would consist of C+AC+BC+ABC.

Each of the crossed factors and their interactions will be tested for significance along with its interaction with the nesting factors and their combination, using a pooled error term consisting of the interaction between each combination to be tested and the nested factor. Thus if the first crossed factor is D, then D, AD, BD, and ABD would be tested against the pooled error term DC+ADC+BDC+ABDC. When replication is present, the "within cells" mean square is used as denominator to perform F tests on each error term. These are interpreted like the similar tests under "Crossed Designs."

VII. POST HOC COMPARISONS

Comparisons between main-effect means are made according to the methods outlined in Winer (1962, pp. 77-89). These methods have been extended to permit testing the differences between the means for different levels of one factor within each level of a second factor or within each combination of levels of two other factors.

The user may choose between Newman-Keuls tests and the somewhat more conservative Tukey (b) tests*. Results are printed out in the form of a lower triangular matrix showing the value of the q statistic for each combination of levels of the test factor. These are marked with asterisks to indicate significance levels. One such matrix is printed out for each level, or combination of levels, of additional factors, within which the test factor is evaluated.

When differences between levels of the test factor are to be tested overall, the error term used is the same one used in the significance test of that factor in the analysis of variance; when the test is to be made within levels of another factor, the error term is obtained by pooling the error term for the test factor with that for its interaction with the other factor; when the test is within combinations of levels of two other factors, it is obtained by pooling those for the test factor, the interactions between the test factor and each of the other factors, and the three-way interaction. These routines may also be used independently with the results of a previously conducted analysis of variance as input data. The procedure for this is described in the Appendix.

---

*See Petrinovich and Hardyck (1969) for a discussion of error rates with these techniques.

VIII. SAMPLE RUN

The following setup is for analysis of data from a hypothetical experiment employing 80 subjects, 40 male and 40 female. Half the males and half the females are used as drivers in the experiment, the others being used as passengers (the "task" factor).

Note that the number of levels of the subject factor (or, more properly, pseudo-factor) is only 20. In nested designs, the number of levels of the subject factor is equal to the number of subjects used <u>in each combination of the nesting factors</u>.

Each subject's reaction time is measured as he responds to signals given by five different signaling systems. Each system is presented four times in each of four different modes for a total of 80 trials per subject. (Hence the number of replications is four and they are correlated, since they represent repeated measures on the same subjects.)

Each subject's data occupy a total of 40 cards, two for each combination of system and mode. The data are integers and are punched as follows: The first trial for each combination of system and mode is in columns 51-60 of the first card, the second trial in columns 71-80 of the first card, the third trial in columns 11-20 of the second card, and the fourth trial in columns 31-40 of the second card. The remaining columns are assumed to be blank or filled with information irrelevant to this analysis.

Each subject's cards are contiguous in the deck and are sorted in ascending order by system, and, within each system, in ascending order by mode. The subjects are arranged in the following sequence: male drivers, male passengers, female drivers, female passengers. The raw data are to be transformed to natural logarithms.

Post hoc comparisons between systems are desired for each mode. In addition, a four-way table showing means for each combination of the four fixed factors, and a two-way table of means for the combinations of sex and task, are desired.

The input data are stored in the file DATAFILE, followed by some additional data to be used in a second analysis following this one. The analyses are to be made in batch mode, and only the portion of the deck preceding the entry of parameters for the second analysis is shown.

The set-up cards required are shown in Figure 1, on the following page.

HSRI

HSRI Form 31

PROGRAM SAMPLE SETUP-HUMAN FACTORS ANOVA PROG.

PROGRAMMER J. S. L. DATE 7/16/71

NOTE: WHERE A CARD CONTAINS ONLY ZEROS, A BLANK CARD MAY BE USED.

PAGE 1 OF 1

```
    $RUN SAYM:ANØVA 2=DATAFILE 3=*SINK* 4=*SØURCE* 5=*DUMMY*
1.  SAMPLE RUN ØF HUMAN FACTØRS ANØVA PRØGRAM WITH PØST HØC CØMPARISØNS  <-- Label
2.   5 4 4 MØDE 5 SYST20 SUBJ 2 TASK 2 SEX  <-- Design
3.  02      <-- Replications Correlated.
4.  01      <-- Random Factor
5.  C D E   <-- C nested under D & E
6.  0       <-- Normal Data Sort
7.  00004   <-- 4 observations per data group
8.  (50X,F10.0,10X,F10.0/2(10X,F10.0))  <-- format of above group
9.  2       <-- Transform data to nat. logs.
10. 00000   <-- Do not add a constant
11. 02      <-- Do not print cell means
12.{01      <-- Do a post hoc comparison
   {B A     <-- Compare levels of B within each level of A
   {02      <-- Tukey (b) test desired
   {02      <-- End comparisons
13.{01      <-- Print a table
   {B A C D <-- Cols. = systems; rows = modes; small blocks = tasks; large blocks = sexes
   {01      <-- Print a table
   {C D     <-- Cols. = tasks; rows = sexes
   {02      <-- End of tables
14. 01      <-- Begin new analysis
15. 02      <-- Do not rewind (new data follows orig. data in DATAFILE)
16. TEST ØF RERUN FACILITY  <-- Label for second analysis
17. 02      <-- New cell design - return to #3 & follow with rest of setup for second anal.
```

Figure 1. Setup cards needed for sample problem.

REFERENCES

Bogartz, W. "CODE": A Calculus for the Orthogonally Designed Experiments. *Psychological Bulletin*, 1968, *69*, 418-422.

Clifford, T. Simplified Computational Programming for Analysis-of Variance Designs with Correlated Observations. *Psychological Bulletin*, 1968, *69*, 439-440.

Hartley, H. Analysis of Variance. Chap. 20, *Mathematical Methods for Digital Computers*. Ralston, A. and Wilf, H. (Eds.) New York: John Wiley & Sons, 1962, 221-230.

McNemar, Q. *Psychological Statistics*. New York: John Wiley & Sons, 1962.

Petrinovich, L. and Hardyck, C. Error Rates for Multiple Comparison Methods: Some Evidence Concerning the Frequency of Erroneous Conclusions. *Psychological Bulletin*, 1969, *71*, 43-54.

Winer, B. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1962.

Appendix I

Program to Perform *Post Hoc* Comparisons On Results of Previous Analyses of Variance

This program permits the user to utilize the *post hoc* comparison routines from the HSRI Human Factors ANOVA program independently to compare means from analysis of variance, regardless of whether it was performed by the ANOVA program.

The means to be compared, the appropriate error mean square, and the proper number of degrees of freedom must be supplied by the user.

RUNNING THE PROGRAM

The program is initiated by the command

$RUN SAYM:POSTHOC 3=BULKFILE 4=*SOURCE* 5=*SINK*
for terminal use. For batch use I/O unit #3 should be assigned to *SINK* and #5 to *DUMMY*.

The set-up is as follows:

1. Label Card. Up to 80 characters to appear as a heading on the printed output.

2. Parameter Card.

Col.	1- 5	Total number of observations in the analysis.
Col.	6-10	Number of levels of factor to be tested.
Col.	11-15	Number of levels of first factor to be held constant.
Col.	16-20	Number of levels of second factor to be held constant.
Col.	22-25	Four-character name of factor to be tested.
Col.	26-30	Name of first factor to be held constant.
Col.	31-35	Name of second factor to be held constant.
Col.	40	0 if table is wanted*. 1 if no table wanted.

*The table referred to is a lower triangular matrix containing the values of the *q* statistic and one asterisk for values significant at the .05 level, or two asterisks for those significant at the .01 level. If this table is not requested, the values are printed, one per line, at the left of the paper.

Col. 45 0 for Newman-Keuls.
1 for Tukey (b).

3. Error Term Card*.

Col. 0-10 Error mean square for main effect of tested factor.

Col. 11-20 Error mean square for interaction between tested factor and first factor to be held constant (if any).

Col. 21-30 Error mean square for interaction between tested factor and second factor to be held constant (if any).

Col. 31-40 Error mean square for 3-way interaction of above factors (if any).

Note: Unless the decimal point is explicitly punched, it is assumed to be between columns 5-6, 15-16, etc. (F10.5 format).

4. Degrees of Freedom Card. Should contain the number of degrees of freedom associated with each of the above error terms.

Col. 1- 4 d.f. for first error term.

Col. 5- 8 d.f. for second error term (if any).

Col. 9-12 d.f. for third error term (if any).

Col. 13-16 d.f. for fourth error term (if any).

In all-fixed-factor designs where all of the above main effects and interactions are tested against the same error term, this error term and its d.f. should be entered only once.

5. Data Deck. The means to be tested should be punched eight to a card, in 8-column fields starting in column 1. The decimal point is assumed to be between the fifth and sixth columns of the field unless explicitly punched (F8.3 format). The means should be in the following order:

If there are two constant factors, starting with level 1 of each constant factor the means for each level of the tested factor should occur in natural order. Repeat for level 2 of the

---

*The "error mean square" for a given factor or interaction refers to the error term which was used as the denominator of the F ratio for that factor or interaction in the analysis of variance.

first constant factor, and so on until all its levels have been exhausted. Then proceed to level 2 of the second constant factor and repeat the previous procedure. Continue in this fashion until all the means have been punched.

For example, if $\bar{X}_{\underline{ijk}}$ is the mean for the $\underline{i}^{th}$ level of the first constant factor, and the $\underline{k}^{th}$ level of the second constant factor, and if each factor has three levels, the means would be punched as follows:

Card 1. $\bar{X}_{111}$, $\bar{X}_{211}$, $\bar{X}_{311}$, $\bar{X}_{121}$, $\bar{X}_{221}$, $\bar{X}_{321}$, $\bar{X}_{131}$, $\bar{X}_{231}$

Card 2. $\bar{X}_{331}$, $\bar{X}_{112}$, $\bar{X}_{212}$, $\bar{X}_{312}$, $\bar{X}_{122}$, $\bar{X}_{222}$, $\bar{X}_{322}$, $\bar{X}_{132}$

Card 3. $\bar{X}_{232}$, $\bar{X}_{332}$, $\bar{X}_{113}$, $\bar{X}_{213}$, $\bar{X}_{313}$, $\bar{X}_{123}$, $\bar{X}_{223}$, $\bar{X}_{323}$

Card 4. $\bar{X}_{133}$, $\bar{X}_{233}$, $\bar{X}_{333}$

6. Terminating the Run. If another set of comparisons is wanted, the data deck should be followed immediately by another setup, starting with a new label card (item 1). If no more analyses are to follow, the data should be followed by a $ENDFILE card (or by entering an end-of-file character, if running from a terminal).

Appendix II

SUMMARY OF ANOVA SET-UP CARDS

	CARD		REQUIRED IF:
	1.	Title or header	Always on 1st anal.
	2.	Cell design	Always on 1st anal.
	3.	Correlated replication?	>1 score/cell
	4.	Random factor? (yes/no) (Note 2)	Always
	5.	Random & nesting factors	Previous card = 01 (yes)
	6.	Type of data sort	Cell freq. equal and >1
	7.	Observations per data group	Always
	8.	Data format	Always
	9.	Transformation*	Always
	10.	Augmenting constant	Transform. other than arc sine (prev. card = 1, 2, or 3)
	11.	Events/observ; Freq. or proportions	Arc sine trans.
	DATA (IF SUBMITTED WITH SETUP)†		
	12.	Cell means	Always
Note 1	13.	Do multiple comparisons?	Always
Note 1	14.	Mult. comp. factor card	Mult comp. wanted
Note 1	15.	01 for Newman-Keuls, 02 for Tukey (b)	Mult comp. wanted
Note 1	16.	2-, 3-, 4-way table?	Always
Note 1	17.	Table setup card	Table wanted
	18.	Do another ANOVA?	Always
	19.	Rewind data set?	Another ANOVA wanted
	20.	Title or header	Another ANOVA wanted
	21.	Same cell design? (If yes go back to 3; If no go to 2.)	Another ANOVA wanted

---

Note 1: These groups are repeated for as many tests or tables as wanted. The last one must be followed by an "02" card to end the series.

Note 2: All yes-no cards use 01=yes, 02=no.

*0=NONE 1=SQUARE ROOT 2=LOG_e 3=RECIPROCAL 4=ARC SINE

†If *n* unequal, card 1=Max. cell *n*, cards with cell *n*'s (20/card) must precede data.

TABLE 1. SUMMARY OF ANOVA SET-UP CARDS

	CARD		REQUIRED IF:
	1.	Title or header	Always on 1st anal.
	2.	Cell design	Always on 1st anal.
	3.	Correlated replication?	>1 score/cell
	4.	Random factor? (yes/no) (Note 2)	Always
	5.	Random & nesting factors	Previous card = 01 (yes)
	6.	Type of data sort	Cell freq. equal and >1
	7.	Observations per data group	Always
	8.	Data format	Always
	9.	Transformation*	Always
	10.	Augmenting constant	Transform. other than arc sine (prev. card = 1, 2, or 3)
	11.	Events/observ; Freq. or proportions	Arc sine trans.
	DATA (IF SUBMITTED WITH SETUP)†		
	12.	Cell means	Always
Note 1	13.	Do multiple comparisons?	Always
Note 1	14.	Mult. comp. factor card	Mult comp. wanted
Note 1	15.	01 for Newman-Keuls, 02 for Tukey (b)	Mult comp. wanted
Note 1	16.	2-, 3-, 4-way table?	Always
Note 1	17.	Table setup card	Table wanted
	18.	Do another ANOVA?	Always
	19.	Rewind data set?	Another ANOVA wanted
	20.	Title or header	Another ANOVA wanted
	21.	Same cell design? (If yes go back to 3; If no go to 2.)	Another ANOVA wanted

Note 1: These groups are repeated for as many tests or tables as wanted. The last one must be followed by an "02" card to end the series.

Note 2: All yes-no cards use 01=yes, 02=no.

*0=NONE 1=SQUARE ROOT 2=LOG_e 3=RECIPROCAL 4=ARC SINE

†If n unequal, card 1=Max. cell n, cards with cell n's (20/card) must precede data.

The above table (Table 1) briefly summarizes the setup cards necessary to run ANOVA. It has been reduced so that it can be detached and mounted on a 5" x 8" card for convenient use as a quick reference.