HSD ANOVA


A GENERAL ANALYSIS OF VARIANCE PROGRAM

FOR THE APPLE II



by

Stephen Madigan, Ph.D.

and

Virginia Lawrence, Ph.D.

# H S D    A N O V A

### A GENERAL ANALYSIS OF VARIANCE PROGRAM
### FOR THE APPLE II*

HSD ANOVA provides a flexible, general and easy to use analysis of variance procedure, with these features:

Analysis of any balanced design, from a simple two-group design to complex factorial designs with up to 8 variables.

Analysis of designs with between-subjects and/or within-subjects factors, with simple design specifications and data entry.

Production of a fully formatted ANOVA summary table showing: experiment name; experimental effects names; sums of squares; degrees of freedom; mean squares; F-ratios; F probabilities; means and standard deviations.

| SOURCE | SS | DF | MS | F | P |
|--------|--------|----|--------|--------|------|
| AAA | 15.042 | 1 | 15.042 | 7.847 | .012 |
| BBB | 1.792 | 3 | .597 | .311 | |
| AAA BBB | 149.458 | 3 | 49.819 | 25.988 | .000 |
| ERROR | 30.667 | 16 | 1.917 | | |

Program options include:

- user specification of variable names and condition names.
- output to TV monitor or 40- or 80-column printer.
- data entry from keyboard or from disk.
- data output to disk.
- data review and editing at input.

### SYSTEM REQUIREMENTS

Apple II or Apple II+ with 48K, disk drive and ROM APPLESOFT.
A 48K Apple II can handle a design with 1,200 observations.

--------------------

*Apple II is a registered trademark of Apple Computer, Inc.

# I N S T R U C T I O N S

HSD ANOVA will perform an analysis of the variance for most standard experimental designs consisting of from 1 to 8 factors. The number of levels of variables and the number of observations are limited only by the memory limitations of your system. The program requires "equal n" for all designs. Fixed effects are assumed for all factors in the calculation of F-ratios. However, F-ratios for random effects can be obtained after calculating the appropriate error terms from the sums of squares produced by the program.

The use of HSD ANOVA will be illustrated by going through a simple problem step by step. Suppose you have an experimental design with 3 independent variables or factors: Factor A has 2 levels, Factor B has 2 levels, and Factor C has 3 levels. There are 5 scores (observations or data points) in each of the 12 (2 x 2 x 3) conditions, as shown in the following table.

|     |     | b1 |     |     | b2 |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|
|     |     | c1 | c2 | c3 | c1 | c2 | c3 |
| a1  |     | 6 | 8 | 13 | 3 | 3 | 4 |
|     |     | 9 | 7 | 11 | 5 | 3 | 5 |
|     |     | 10 | 12 | 10 | 3 | 6 | 3 |
|     |     | 12 | 9 | 16 | 7 | 2 | 2 |
|     |     | 8 | 12 | 10 | 2 | 4 | 4 |
| a2  |     | 7 | 5 | 10 | 10 | 11 | 5 |
|     |     | 9 | 13 | 12 | 9 | 9 | 6 |
|     |     | 8 | 8 | 10 | 8 | 14 | 8 |
|     |     | 12 | 9 | 11 | 9 | 14 | 9 |
|     |     | 12 | 11 | 14 | 12 | 9 | 12 |

Assume also that Factor A is a between-subjects factor, and Factors B and C are within-subjects factors. The analysis of these data would proceed as follows:

1.   Boot the HSD ANOVA disk.  The HSD title will appear.  RETURN to continue.
     (In all of the subsequent steps, keyboard entry of an improper or illegal
     quantity in response to a request for input will cause a "bell" to sound
     twice, followed by a repeat of the input prompt.  "YES" or "Y", "NO" or "N"
     are accepted in response to queries.)

2.   The program requests a descriptive name for the analysis.  The name can be
     a string of up to 40 characters and it is printed with the results.

3.   The program asks you to specify the output method, type of printer and
     printer slot number.  In its present form, the program package will operate
     through the Apple Serial Interface, the Silentype/Trendcom Interface, and
     the Centronics Parallel Interface.

```
--------------------
--------------------
SPECIFY OUTPUT METHOD
TV MONITOR      = 1
40-COL. PRINTER = 2
80-COL. PRINTER = 3
ENTER 1, 2 or 3 : 3
SERIAL (S) OR PARALLEL (P)? S
PRINTER SLOT # ? 1
--------------------
--------------------
```

     NOTE:  Only the 80-column option will produce an ANOVA table in its
     standard form, due to line length restrictions with the other two options.
     With a 40-column printer and with the TV monitor output, the lines of the
     ANOVA table are printer vertically.  This is shown in a later example.

4.   The program then asks you to enter the number of factors in the
     experimental design:

```
--------------------
--------------------
NUMBER OF FACTORS : 3
--------------------
--------------------
```

     (An out-of memory error can occur here if the design is too large for your
     system.)

5.  The program then asks you to specify the experimental design, factor by
    factor:

```
                    ---------------------
                    ---------------------
                        FACTOR NUMBER 1
                     NAME OF FACTOR : AAA
                     NUMBER OF LEVELS?  2
                 TYPE OF FACTOR (B=1, W=2) ? 1
                 ENTER LEVEL NAMES, 1 PER LINE
                             A1
                             A2



                        FACTOR NUMBER 2
                     NAME OF FACTOR : BIT
                     NUMBER OF LEVELS?  2
                 TYPE OF FACTOR (B=1, W=2) ? 2
                 ENTER LEVEL NAMES, 1 PER LINE
                             B1
                             B2



                        FACTOR NUMBER #
                     NAME OF FACTOR?  CCO
                     NUMBER OF LEVELS?  3
                 TYPE OF FACTOR (B=1, W=2) ? 2
                 ENTER LEVEL NAMES, 1 PER LINE
                             C1
                             C2
                             C3
                    ---------------------
                    ---------------------
```

- The name of the factor can be up to 3 characters long. The first character of every name must be a letter. Names which are longer than 3 letters will be truncated to 3.

- The levels of every factor must be 2 or more. (Out-of-memory may occur here too.)

- The type of factor is defined by the user by entering 1 for a between-subjects factor, or a 2 for a within-subjects factor. No further specification of the type of experimental design is needed by HSD ANOVA.

- The names of the levels of each factor can be up to 3 characters long.

NOTE: The order in which the factors are entered is critical. See Section 7 on data input order.

6. The program now requests the number of scores per condition. Two or more scores are required per condition.

```
---------------------
---------------------
NUMBER OF SCORES PER CONDITION : 5
---------------------
---------------------
```

7.A. DATA ENTRY
   The program asks if you want data retrieved from disk (sequential text file
   only). A YES response leads to a request for the file name; the file is
   retrieved as soon as a RETURN occurs. Since only one drive can be used
   with this program, the data will be read by the same drive used to read the
   program. The length of the file must be greater than or equal to the
   number of scores specified by the design (60 in the present example). If
   the length of the file is less than the number of scores in the design, or
   if the file is not on the disk, an end-of data message will occur and
   execution will halt. Naming a non-existent file will cause that file name
   to be added to your disk catalog. You can executive DELETE (filename) to
   remove it.

```
                  ---------------------
                  ---------------------
                       DATA ENTRY
                READ DATA FROM DISK ? NO
                  ---------------------
                  ---------------------
```

   The NO response leads to a request for keyboard entry of the data for the
   first cell or condition of the experiment design. The program prints the
   observation number and waits for the appropriate number entry followed by
   RETURN .

```
                  ---------------------
                  ---------------------
            DATA FOR CONDITION A1, B1, C1
                       1.   6
                       2.   9
                       3.  10
                       4.  12
                       5.   8

            DATA FOR CONDITION A1, B1, C2
                       1.   8
                       2.   7
                       3.  12
                       4.   9
                       5.  12

                        etc.
                  ---------------------
                  ---------------------
```

   The screen will accept up to 30 scores at a time, in 3 columns of 10. If
   there are more than 30 to be entered for a single condition, the screen
   clears and resets before continuing.

Data entry continues until all design conditions have been entered. There is a pause after the entry of all data for a condition. The "bell" sounds once when the computer is ready to accept data for the next condition.

7.B. ORDER OF DATA ENTRY

There are 2 requirements for the order of data entry which must be followed for a valid analysis of variance.

(i) Data are entered with the last subscript or dimension of the data array varying first (or fastest). In the current example, data are entered in the order a1b1c1, a1b1c2, a1b1c3, .... a2b2c3. All data for a given condition are entered before moving to the next condition. This rule also applies to data read from disk. Data read from disk must have been previously stored in this order. For the example, the file would have to contain the numbers 6, 9, 10, 12, 8, 8, 7, 12, ...., 5, 6, 8, 9, 12 in that order.

(ii) The array of data must be structured so that between subjects factors (if any) occur first, and within-subjects factors (if any) occur last. This is illustrated in the example and is also the common practice followed in statistics and research methods texts.

8. DATA STORAGE

After the ANOVA data entry, you are asked whether you want your data stored on disk. If you want to store your data, you will be asked to specify a file name for storage. If you have edited data previously read from disk, use of the same file name will result in the changes being stored and written over the original data. After the analysis is completed, you will be asked again whether you want to write data to disk.

9. THE ANALYSIS

As soon as the data from the last condition are entered, the analysis proceeds, first producing a listing of the mean and standard deviation of each cell of the design. Each cell is identified by its condition names. When data are read from disk, CONDITION and NAME will be displayed while means and standard deviations are being computed.

```
------------------------
------------------------
CONDITION      MEAN      SD
------------------------------
A1 B1 C1       9.000     2.236
A1 B1 C2       9.600     2.302
A1 B1 C3      12.000     2.550
A1 B2 C1       4.000     2.000
                etc.
------------------------
------------------------
```

With TV monitor output, these results appear on the screen as shown here. Up to 8 will be displayed at one time. For more than 8, a RETURN FOR MORE DISPLAY message will appear.

When means and standard deviations for all conditions have been printed, the program moves into the ANOVA routine. While calculations are proceeding, the TV screen will indicate which experimental effect is being computed by displaying the name of the effect. This will inform you that the program is still computing and is not hung up somewhere. Effect names combined with SSS are displayed when error terms are being computed.

When the analysis is finished, the ANOVA table appears. With an 80-column printer, it looks like this:

| SOURCE | SS | DF | MS | F | P |
|---|---|---|---|---|---|
| AAA | 126.150 | 1 | 126.150 | 17.321 | .003 |
| ERROR | 58.267 | 8 | 7.283 | | |
| | | | | | |
| BIT | 176.817 | 1 | 176.817 | 38.581 | .000 |
| AAA BIT | 138.016 | 1 | 138.016 | 30.115 | .001 |
| ERROR | 36.667 | 8 | 4.583 | | |
| | | | | | |
| CCO | 4.933 | 2 | 2.467 | .510 | |
| AAA CCO | 8.400 | 2 | 4.200 | .869 | |
| ERROR | 77.333 | 16 | 4.833 | | |
| | | | | | |
| BIT CCO | 46.533 | 2 | 23.267 | 6.317 | .010 |
| AAA BIT CCO | 8.534 | 2 | 4.267 | 1.159 | .340 |
| ERROR | 58.933 | 16 | 3.683 | | |

With a 40-column printer or TV monitor output, it appears as follows:

```
----------------------
----------------------
AAA
----------------------
SS              126.150
DF                    1
MS              126.150
F                17.321
P                  .003
----------------------

ERROR
----------------------
SS               58.267
DF                    8
MS                7.283
----------------------

BIT
----------------------
SS              176.817
DF                    1
MS              176.817
F                38.581
P                  .000
----------------------

           etc.
----------------------
----------------------
```

With TV monitor output, you will be prompted with  RETURN  to obtain the display of successive effects.

10.  DATA HARD COPY

If you are using an 80-column printer, you can obtain a copy of the data entered by replying to the hard copy query.  The data are printed by rows in the order entered, along with the assigned condition names.

11.  DATA EDITING

As soon as you have finished data entry for a given condition, a DATA
REVIEW/EDIT prompt appears.  If you respond YES, you can review and/or edit
your data condition by condition.

The data re first displayed, then you are asked if you want to edit.  If
you reply YES, you are asked to specify which data point you want to change
by entering the number of the score, then the new value.

The data display and editing routine will show up to 10 scores at a time on
the TV monitor, and allow editing of the displayed scores.  If there are
more than 10 scores per condition, the program will continue to display
them in 10-score blocks until finished.

You can enter  ?  to end editing of data from the current condition.

Data from the first condition of the example will be used to illustrate
keyboard entry data editing:

```
                    ---------------------
                    ---------------------
          REVIEW DATA FOR CONDITION A1 B1 C1 ? Y
                    CONDITION A1 B1 C1
                      ----------------
                        1.   1
                        2.   7
                        3.   9
                        4.  11
                        5.  12
                      ----------------
                      EDIT ANY OF THESE DATA?  Y
                      ENTRY #? 1
                      NEW VALUE? 6
                      ANY MORE CHANGES?  Y
                      ENTRY #?  2
                      NEW VALUE?  9
                      ANY MORE CHANGES?  N
                    ---------------------
                    ---------------------
```

## ADDITIONAL PROGRAM INFORMATION

1. Probability values. In the ANOVA table, p-values are computed for any F-ratio greater than 1.0. A probability less than .001 appears in the table as .000, but should be read as "probability less than .001."

2. Number format. The ANOVA table is formatted to allow display of numbers to 3 places to the right of the decimal point and six places to the left of the decimal point. Numbers greater than 999999.999 cannot be accommodated because of formatting problems. If such numbers are encountered, the program will halt. In such a case, reduce the absolute magnitude of your data. This can be done by dividing all data points by a constant, such as 10. The maximum printable values for F and df are 9999.999 and 9999, respectively.

3. Accuracy. The program has been tested against many example problems in several standard experimental design and statistics texts. The program has also been tested against solutions produced by well-known statistical packages run on large systems.

4. Execution Time. With keyboard entry of data, means and standard deviations will appear immediately following entry of data for the last condition. With disk entry of the data, there will be a pause before the results appear. The main execution time (from the end of the listing of means and standard deviations to the appearance of the first line of the ANOVA table) depends greatly on the kind of experimental design being analyzed. It is influenced mostly by the number of factors in the design, and by the number of factors which are within-subjects factors. An exception is the single-factor design where the execution time is approximately the same whether the factor is a between- or within-subjects factor. The number of scores per condition does not have much influence on execution times for ramdomized designs, but does have a very large influence on designs with within-subjects factors.

   Some examples of execution times for various designs are given below:

   | DESIGN | FACTORS | N/CONDITION | TIME (sec.) |
   |---|---|---|---|
   | 1 x 5 | B | 10 | 3 |
   | 1 x 10 | B | 20 | 4 |
   | 2 x 2 | B,B | 10 | 5 |
   | 2 x 2 | W,W | 10 | 12 |
   | 2 x 2 x 2 | B,B,B | 20 | 14 |
   | 2 x 2 x 2 | W,W,W | 20 | 70 |
   | 2 x 2 x 5 | B,W,W | 10 | 70 |
   | 2 x 2 x 2 x 2 | B,B,B,B | 10 | 43 |

5.  Hierarchical Designs.  The program can be used to analyze at least some
    kinds of hierarchical designs.  This is accomplished by entering the design
    as if all variables were crossed.  When the ANOVA table has been obtained,
    the sums of squares can be recombined to obtain the appropriate error
    terms.

6.  Random Effects.  The program computes F-ratios for fixed effects only.
    However, the F-ratios for random effects can be computed from the SS and df
    terms produced by the program.

-------------------------------------------------------------------------------

Any inquiries can be sent to:
    HUMAN SYSTEMS DYNAMICS
    Suite 107
    9249 Reseda Boulevard
    Northridge, CA 91324
    (213) 993-8536

# Example

ENTER NAME OF PROBLEM:  2-FACTOR DISK INPUT

SPECIFY OUTPUT METHOD


TV MONITOR      = 1

40-COL. PRINTER = 2

80-COL. PRINTER = 3


ENTER 1, 2 OR 3:  3

SERIAL (S) OR PARALLEL (P):  S

PRINTER SLOT #: 1
NUMBER OF FACTORS:  2

FACTOR NUMBER:  1
------------------

NAME OF FACTOR:  AAA

NUMBER OF LEVELS:  2

TYPE OF FACTOR (B=1, W=2)  1

ENTER LEVEL NAMES, 1 PER LINE

LEVEL 1  A1

LEVEL 2  A2


FACTOR NUMBER 2
----------------

NAME OF FACTOR:  BBB

NUMBER OF LEVELS:  2

TYPE OF FACTOR (B=1, W=2)  1

ENTER LEVEL NAMES, 1 PER LINE

LEVEL 1 B1

LEVEL 2 B2

NUMBER OF SCORES PER CONDITION:  5

```
** DATA ENTRY **
READ DATA FROM DISK?  YES

FILE NAME:  FV145
```

CONDITION A1  B1


```
REVIEW/EDIT DATA?  N
```
CONDITION A1  B2


```
REVIEW/EDIT DATA?  N
```
CONDITION A2  B1


```
REVIEW/EDIT DATA?  N
```
CONDITION A2  B2


```
REVIEW/EDIT DATA?  N
```

2-FACTOR DISK INPUT


PROBLEM DESCRIPTION

| FACTOR | NAME | LEVELS | TYPE |
|--------|------|--------|------|
| 1 | AAA | 2 | B |
| 2 | BBB | 2 | B |

N PER CONDITION:  5

INPUT FILE NAME:  FV145

| CONDITION | | MEAN | S.D. |
|-----------|-----|--------|--------|
| A1 | B1 | 54.400 | 19.269 |
| A1 | B2 | 56.600 | 8.264 |
| A2 | B1 | 60.600 | 8.849 |
| A2 | B2 | 51.000 | 11.000 |

| SOURCE | SS | DF | MS | F | P |
|--------|-----|-----|-----|-----|-----|
| TOTAL | 2798.550 | 19 | | | |
| AAA | .450 | 1 | .450 | .003 | |
| BBB | 68.450 | 1 | 68.450 | .429 | |
| AAA BBB | 174.050 | 1 | 174.050 | 1.090 | .313 |
| ERROR | 2555.600 | 16 | 159.725 | | |