DECUS

PROGRAM LIBRARY

DECUS NO.	12-38A
TITLE	HISTOGRAM AND ONE-FACTOR ANALYSIS OF VARIANCE
AUTHOR	Mary Kathleen Fairbanks
COMPANY	Neuropsychology Research Veterans Administration Hospital Sepulveda, California
DATE	August 5, 1971
SOURCE LANGUAGE	FOCAL-12

HISTOGRAM AND ONE-FACTOR ANALYSIS OF VARIANCE

<u>DECUS Program Library Write-up</u> <u>DECUS NO. 12-38A</u>

Hardware Requirements: PDP-12A, two TU55 tape units, 8K memory.

Language: Focal-12

The Program: Performs three primary functions which may be executed singly or in any desired combination i.e., data storage, histogram construction, and analysis of variance computation. Accepts integer data entered via teletype and stores these data on LINC tape using the Dial index. Displays a histogram of the integers on request using the PDP-12 scope. Displays minimum, second smallest, second largest and maximum values of the data array. Computes either a one-factor repeated measures or a one factor completely randomized analysis of variance on the data if requested. This program package is composed of the following program segments: $ANOVA, $HISTGM, $INT, $GPH, %AV, %2AV, %3AV. The package will handle a maximum of 6ØØ numbers at one time and the largest number of intervals that the histogram may have is 95.

Operation: Place the LINC tape containing Focal-12 and the program package on unit Ø. Place a second LINC tape containing Dial on unit 1. Call Dial, load Focal-12, then type "L G,$ANOVA,Ø". The space bar is the terminal character for all lines in the program package. It should be typed after responding to each question in the programs. For any questions requiring a yes or no answer, the correct response is either "Y" (yes), or "N" (no).

<u>$ANOVA</u>:

HOW MANY TRIALS?: (response varies according to design of experiment)
a) for data from a one-factor <u>repeated measures</u> design (c.f. Meyers, <u>Fundamentals of Experimental Design</u>, p. 152), respond with the number of levels of the treatment variable (A) i.e., the number of trials given to a subject. It is necessary that this number be equal for all subjects.
b) for data from a one-factor <u>completely randomized</u> design (c.f. Meyers, p. 51), respond with the number of levels of the treatment variable A.
c) for <u>all other</u> designs, respond with a number that is a factor of the total number of data points to be entered i.e., if 5Ø numbers are to be entered, the response could be either 1Ø or 25, then the answer to the next question would be either 5 or 2. The product of the numbers given as answers to the first two questions must equal the total number of data points to be entered.

HOW MANY SUBJECTS?: (response depends upon experimental design)
a) for a one-factor <u>repeated measures</u> design, respond with the total number of subjects (S) tested.
b) for a <u>completely randomized</u> one-factor design, respond with the number

$ANOVA (con't):
of subjects tested at one level of the treatment variable i.e., the number of subjects in one group.
c) for all other designs, respond with that number which when multiplied by the number given in response to the first question will equal the total number of data points to be entered on tape or brought down from storage.
ARE DATA ON TAPE?:
"Y" response causes computer to locate and display on the scope data previously stored on LINC tape. If there are too many data points to display simultaneously, the scope will fill up with the first segment of data and display this until the operator has checked the entries and then types line feed to advance the display to the next segment of the data. If any errors are discovered in the data list at this time, the ordinal position of these entries should be recorded so that later modification may be made.
"N" response will make and open a 10 block data file on unit 1. The scope will clear and a colon (:) will appear on the screen after which the first number in the data array should be typed followed by the terminal character. Colons will continue to appear on the scope as long as more data are to be entered. (Remember to type the terminator after each entry.)

The order in which the data are entered is determined by the experimental design.
a) one-factor repeated measures design - enter the data from one subject at a time, from first score to last score. Assume 2 subjects (S) are tested three times each (A represents the trial numbers).

A_1	A_2	A_3
S_1	S_1	S_1
S_2	S_2	S_2

These six scores must be entered in this order:
first: S_1A_1
second: S_1A_2
third: S_1A_3
fourth: S_2A_1
fifth: S_2A_2
sixth: S_2A_3
b) one-factor completely randomized design -- enter the data from the first subjects in each of the a treatment groups, then from the second subjects in each group, and so on until the scores from the n^{th} subjects in each group have been entered. Assume 6 subjects (S) are randomly assigned to the three levels of the treatment variable (A). There will be three treatment groups with two different subjects in each.

A_1	A_2	A_3
S_1	S_3	S_5
S_2	S_4	S_6

These data are to be entered in this order:
first: S_1A_1
second: S_3A_2
third: S_5A_3
fourth: S_2A_1
fifth: S_4A_2
sixth: S_6A_3

$ANOVA (con't):
c) all other designs -- the order of entry is immaterial.

After all data have been entered they will be displayed on the scope so that they may be checked for errors during entry as described above.

TYPE "C" TO CONTINUE: Indicates the end of the data list. "C" response will cause the program to continue.

MODIFY DATA?:
"N" response causes computer to by-pass the correction sequence and begin computing the minimum, maximum, second largest, and second smallest values in the data array.
"Y" response initiates the correction process.

GIVE NO. OF INCORRECT ENTRY: respond with the ordinal position (x) of the incorrect entry in the data list.

FØ(X)=__________The existing value of the X^{th} number in the data array (FØ) is displayed.

NEW VALUE=: respond with the correct value of FØ(X).

MODIFICATION COMPLETE?:
"N" response indicates that more corrections are to be made and displays, GIVE NO. OF INCORRECT ENTRY: again, which begins the correction sequence.
"Y" response will display the data array on the scope once more for a final check. After typing "C" to continue, a response of "N" to MODIFY DATA?: causes the program to go on with a corrected data list.
MIN=__________the value of the smallest number in the data array.
S2=__________the value of the second smallest number in the data array.
MAX=__________the value of the largest number in the data array.
M2=__________the value of the second largest number in the data array.

WANT A HISTOGRAM?:
"N" response indicates that the data are not to be displayed as a histogram.
"Y" response will initiate the histogram sequence of the program package.
It will take 12 minutes for the computer to generate a histogram from an array of 6ØØ numbers with 95 intervals. With less data and/or fewer intervals, the process is more rapid.

ANALYZE DATA?: This is displayed at this point only if an "N" response was given to WANT A HISTOGRAM?
"N" response will cause the computer to halt.
"Y" response will cause the computer to skip the histogram segments and initiate the first computations for the analysis of variance.

$HSTGM: Initiated by a "Y" response to WANT A HISTOGRAM?:
HISTOGRAM
CHOOSE UPPER BOUNDARY
MAX=__________the value of the largest number in the data array.
NEXT MAX=__________the value of the second largest number in the array.

$HSTGM: (con't)
UB /=MAX the upper boundary selected must not exceed MAX.
UB=: respond with the value selected as the histogram's upper boundary.
CHOOSE LOWER BOUNDARY
MIN=__________the value of the smallest number in the data array.
NEXT MIN=__________the value of the second smallest number in the array.
LB /=MIN the lower boundary selected must not be less than MIN.
LB=: respond with the value selected as the histogram's lower boundary.

HOW MANY INTERVALS?: Enter the number of intervals that the histogram is to contain. Since the program handles only integer data, the programmer is cautioned not to request a number of intervals greater than UB-LB. This will prevent distortion of the histogram which would result from asking how many integers fall in intervals not containing any integers. In any case, the number of intervals must not exceed 95. One or two of the intervals may be overflow intervals depending upon the values selected for UB and LB. That is, if the programmer requests 8∅ intervals and chooses for UB and LB, MAX and MIN respectively, he will have a histogram that consists of 8∅ equal intervals. However, if UB is less than MAX the uppermost interval will be an overflow interval of width MAX-UB. The remaining 79 intervals will be equally spaced between MIN (which=LB in this particular instance) and UB. Similarily, if LB is greater than MIN (and UB=MAX), the lowermost interval will be an overflow interval of width LB-MIN. The remaining 79 intervals will be equally spaced between LB and MAX (UB). If both UB is less than MAX and LB is greater than MIN, there will be an overflow interval at each end of the histogram with 78 equally spaced intervals falling between UB and LB.
If the programmer chooses a lower boundary which is smaller than MIN and/or an upper boundary which is larger than MAX he will be given the message(s):
LB IS TOO SMALL
LB /= (MIN) LB cannot be less than MIN.
LB=: Enter a new value for LB.
and/or:
UB IS TOO LARGE
UB /=(MAX) UB cannot be greater than MAX.
UB=: Enter a new value for UB.
When the values of UB and LB are within the proper range the next program in the package will be initiated automatically.

$INT: Initiated by $HSTGM. This segment functions to compute and display on the scope the uppermost boundaries (the X-coordinates) of the intervals in the histogram. The locations of the X-coordinates are computed by dividing the range between UB and LB by the number of intervals desired (less any overflow intervals). If the display of the values of the X-coordinates fills the scope before all coordinates have been shown, advance the display by typing line feed.
TYPE "C" TO CONTINUE: Indicates the end of the listing of coordinates. "C" response will cause the program to continue and initiate the next program segment.

$GPH: Initiated by $INT. This segment functions to compute the Y-coordinates for the histogram and display the completed graph on the scope. The

$GPH: (con't)
Y-coordinates represent the number of points falling at or below the X values for each interval.
ANALYZE DATA?:
"N" response brings the program to a halt.
"Y" response will initiate the next sequence which computes a standard one-factor analysis of variance.

%AV: This segment is initiated by a "Y" response to the question ANALYZE DATA? in either the $ANOVA or the $GPH program segments. It functions to calculate preliminary sums and sums of squares to be used in the computation of the sum of squares, mean squares and F-ratios for the analysis of variance. The computational formulae employed are:

$$\sum^{S}\sum^{A} F\emptyset^{2} \qquad \sum^{S}\sum^{A} F\emptyset \qquad \sum^{S} F\emptyset$$

Where F∅ is the data array, S is the number of subjects (the total number in a repeated measures design or the number of subjects in one group in a completely randomized design) and A is the number of levels of the treatment variable. These computations are stored on LINC tape for use in the next program segment.
SPECIFY DESIGN
C=COMPLETELY RANDOMIZED
R=REPEATED MEASURES
DESIGN=:
"C" response indicates that the data being analyzed conform to a completely randomized one-factor design. The program segment which completes the analysis of variance for the completely randomized design will be initiated by this response.
"R" response indicates that the data being analyzed conform to a one-factor repeated measures design. This response initiates that program segment which functions to complete the analysis of variance for a one-factor repeated measure design.

%2AV: This program is initiated by an "R" response to DESIGN=: in the preceeding program (%AV). Its function is to complete the analysis of variance for a one-factor repeated measures design. It computes F ratios for both A (treatment) and S (subjects), and then prints out this information in a tabular form listing the source of variance (SV), sums of squares (SS), mean squares (MS), F ratios (F, were appropriate) and degrees of freedom (DF). The formulae employed in the calculations are described in Meyers, p. 155.

%3AV: This program is initiated by a "C" response to DESIGN=: in program segment %AV. Its function is to compute the analysis of variance for a completely randomized one-factor design. It computes the F ratio for the treatment effect (A) and then prints out this information in tabular form listing the sources of variance (SV), sums of squares (SS), mean squares (MS), F ratio (F) and degrees of freedom (DF). The formulae employed in the calculations are described in Meyers, p. 7∅.

```
*L L,$ANOVA,Ø
*W
C FOCAL-12

Ø1.Ø1 O S;A "HOW MANY TRIALS?"NT,!,"HOW MANY SUBJECTS?"NS,!
Ø1.Ø4 S TNS=NT*NS;A "ARE DATA ON TAPE?"ANS;O C
Ø1.12 I (ANS-ØY)2.Ø1,2.Ø5,2.Ø1

Ø2.Ø1 L M,1Ø,D1,1;L O,FØ,I,D1,1;F J=1,TNS;A FØ(J)
Ø2.Ø4 GOTO 2.Ø6
Ø2.Ø5 L O,FØ,I,D1,1
Ø2.Ø6 O C;T !;F J=1,TNS;T J,"  ",FØ(J),!!!!
Ø2.Ø7 A "TYPE 'C' TO CONTINUE"C;O C
Ø2.Ø9 A "MODIFY DATA?"W;O C;I (W-ØY)2.2Ø,2.1Ø,2.2Ø
Ø2.1Ø A "GIVE NO. OF INCORRECT ENTRY"X,!;T "FØ(",X,")=",FØ(X),!
Ø2.11 A "NEW VALUE="Z,!;S FØ(X)=Z;O C
Ø2.12 A "MODIFICATION COMPLETE?"X,!;I (X-ØN)2.Ø6,2.1Ø,2.Ø6
Ø2.2Ø S MAX=FØ(1);S M2=1;S MIN=FØ(1);S S2=2Ø47;F K=2,TNS;D 3
Ø2.24 GOTO 4.Ø1

Ø3.1Ø I (MAX-FØ(K))3.14,3.16,3.16
Ø3.14 S M2=MAX;S MAX=FØ(K);GOTO 3.26
Ø3.16 I (FØ(K)-MIN)3.17,3.22,3.22
Ø3.17 I (M2-MIN)3.18,3.2Ø,3.2Ø
Ø3.18 S M2=FØ(K);GOTO 3.2Ø
Ø3.2Ø S S2=MIN;S MIN=FØ(K);R
Ø3.22 I (M2-FØ(K))3.24,3.26,3.26
Ø3.24 S M2=FØ(K);GOTO 3.26
Ø3.26 I (FØ(K)-S2)3.28,3.28;R
Ø3.28 S S2=FØ(K);R

Ø4.Ø1 T "MIN=",MIN,!,"S2=",S2,!,"MAX=",MAX,!,"M2=",M2,!;L M,2,D2,1
Ø4.Ø9 L O,F1,F,D2,1;S F1(1)=MAX;S F1(2)=M2;S F1(3)=MIN;S F1(4)=S2
Ø4.13 S F1(5)=NS;S F1(6)=NT;S F1(7)=TNS;L C,FØ;L C,F1
Ø4.2Ø O S;A "WANT A HISTOGRAM?"ANS,!;O C;I (ANS-ØY)4.24,4.26,4.24
Ø4.24 A "ANALYZE DATA?"Q;O C;I (Q-ØY)4.28,4.27,4.28
Ø4.26 L G,$HSTGM,Ø
Ø4.27 L G,%AV,Ø
Ø4.28 O T;O C;Q
*
```

```
 L L,$HSTGM,0
*W
C FOCAL-12

03.08 O C;L O,F1,F,D2,1;S MAX=F1(1);S M2=F1(2);S MIN=F1(3);S S2=F1(4)
03.19 O S;T "HISTOGRAM",!,"CHOOSE UPPER BOUNDARY",!,"MAX=",MAX,!
03.21 T "NEXT MAX=",M2,!!
03.22 T "UB</=",MAX,!,"UB=";A UB;O C
03.24 T "CHOOSE LOWER BOUNDARY",!,"MIN=",MIN,!,"NEXT MIN=",S2,!!
03.29 T "LB>/=",MIN,!,"LB=";A LB;O C
03.30 A "HOW MANY INTERVALS?",B;O C;S F1(11)=B;S NBI=B-2;S F1(8)=NBI
03.43 I (UB-MAX)3.52,3.46,3.49
03.46 S NBI=NBI+1;S F1(8)=NBI;GOTO 3.52
03.49 O S;T "UB IS TOO LARGE",!,;DO 3.22
03.50 GOTO 3.43
03.52 I (MIN-LB)3.60,3.55,3.58
03.55 S NBI=NBI+1;S F1(8)=NBI;GOTO 3.60
03.58 O S;T "LB IS TOO SMALL",!,;DO 3.29
03.59 GOTO 3.52
03.60 S IR=UB-LB;S II=IR/NBI;S F1(9)=II
03.63 I (MIN-LB)3.69,3.66;GOTO 3.72
03.66 S X(1)=MIN+II;GOTO 3.72
03.69 S X(1)=LB
03.72 S F1(10)=X(1);L C,F1;L G,$INT,0
*
```

```
 L L,$GPH,0
*W
C FOCAL-12

01.09 O C;L O,F0,I,D1,1;L O,F1,F,D2,1;S B=F1(8)
01.13 S MAX=F1(1);S TNS=F1(7);F K=1,TNS;D 2.47
01.20 GOTO 3.51

02.47 F J=9,B;D 2.48
02.48 I (F0(K)-F1(J))2.49,2.49;R
02.49 S V(J)=V(J)+1;S J=B

03.51 F K=9,B;D 3.58
03.57 GOTO 3.83
03.58 F J=1,(V(K)+1);D 3.60
03.60 O S ;S H=FDIS(F1(K)/MAX,J/TNS)
03.83 L C,F0;L C,F1;A "ANALYZE DATA?"RSP;O T;O C
03.91 I (RSP-0Y)3.92,3.94,3.92
03.92 L C,F0;L C,F1;Q
03.94 L G,%AV,0
*
```

```
 L L,$INT,0
*W
C FOCAL-12

03.60 O C;L O,F1,F,D2,1;S II=F1(9);S MAX=F1(1);S X(1)=F1(10)
03.72 S B=F1(11);F J=2,B;S X(J)=X(J-1)+II
03.75 S X(B)=MAX;O S;F J=1,B;T "X(",J,")=",X(J),!!!!
03.79 A "TYPE 'C' TO CONTINUE"C;GOTO 3.80
03.80 S F1(8)=J+7
03.81 F I=1,B;S F1(8+I)=X(I)
03.83 L C,F1;L G,$GPH,0
*
```

```
*L L,%AV,0
*W
C FOCAL-12

01.15 L O,F0,I,D1,1;L O,F1,F,D2,1
01.20 S NS=F1(5);S NT=F1(6);S K=1;S B=NT
01.25 F L=1,NS;S A=K-1+1;S NT=(L-1)*B+B;D 1.35
01.30 GOTO 1.50
01.35 F K=A,NT;D 1.40
01.40 S DSM=DSM+F0(K);S DQ=DQ+F0(K)↑2;D 1.43
01.43 S ASM(K-(L-1)*B)=ASM(K-(L-1)*B)+F0(K);D 1.45
01.45 S SSM(L)=SSM(L)+F0(K)
01.50 S F1(8)=DSM;S F1(9)=B;S F1(11)=DQ
01.53 F R=12,(B+12);S F1(R)=ASM(R-11)
01.55 F T=(B+13),(NS+13+B);S F1(T)=SSM(T-(B+12))
01.56 O S;T "SPECIFY DESIGN",!!,"C=COMPLETELY RANDOMIZED",!
01.57 T "R=REPEATED MEASURES",!
01.58 A "DESIGN=",W,!;O C;I (W-0C)1.60,1.59,1.60
01.59 L C,F0;L C,F1;L G,%3AV,0
01.60 L C,F0;L C,F1;L G,%2AV,0
*
```

```
 L L,%2AV,0
*W
C FOCAL-12

01.10 L O,F1,F,D2,1;S DSM=F1(8);S B=F1(9);S NS=F1(5);S DQ=F1(11)
01.50 S C=DSM↑2/B*NS;S ADF=B-1;F K=12,(B+12);S AQ=AQ+F1(K)↑2
01.60 S QA=AQ/NS-C;S SDF=NS-1;F L=(B+13),(NS+13+B);S SQ=SQ+F1(L)↑2
01.70 S QS=SQ/B-C;S EDF=(NS-1)*(B-1)
01.80 S QD=DQ-C-QA-QS;S AMS=QA/ADF;S SMS=QS/SDF
01.85 S EMS=QD/EDF;S AF=AMS/EMS;S SF=SMS/EMS;O T;T !!!
01.90 T %6.3,"     SV          SS          MS           F            DF",!!
01.92 T "     S        ",QS,"     ",SMS,"    ",SF,"     ",SDF,!
01.94 T "     A        ",QA,"     ",AMS,"    ",AF,"     ",ADF,!
01.96 T "     SA       ",QD,"     ",EMS,"                  ",EDF,!!
01.97 L C,F1;Q
*
```

```
 L L,%3AV,0
*W
C FOCAL-12

01.10 L O,F1,F,D2,1;S DSM=F1(8);S B=F1(9);S NS=F1(5);S DQ=F1(11)
01.50 S C=DSM↑2/B*NS;S ADF=B-1;F K=12,(B+12);S AQ=AQ+F1(K)↑2
01.60 S QA=AQ/NS-C;S TDF=B*NS-1;S QT=DQ-C;S EDF=B*(NS-1)
01.70 S QD=QT-QA;S AMS=QA/ADF;S EMS=QD/EDF;S AF=AMS/EMS;O T;T !!!
01.90 T %6.3,"    SV          SS          MS           F          DF",!!
01.92 T "  TOTAL   ",QT,"                                ",TDF,!
01.94 T "    A     ",QA,"   ",AMS,"   ",AF,"   ",ADF,!
01.96 T "   S/A    ",QD,"   ",EMS,"                      ",EDF,!!;L C,F1;0
*
```