DECUS
PROGRAM LIBRARY

DECUS NO.	FOCAL8-115
TITLE	SHORT PROGRAMS FOR STATISTICAL ANALYSIS USING FOCAL
AUTHOR	D. J. Dowsett and R. Priest
COMPANY	Atkinson Morley's Hospital Wimbledon, England
DATE	February 12, 1970
SOURCE LANGUAGE	FOCAL

SHORT PROGRAMS FOR STATISTICAL ANALYSIS USING FOCAL

DECUS Program Library Write-up

DECUS No. FOCAL8-115

The simple and versatile nature of FOCAL as a programming language has enabled a useful library of statistical routines to be developed for use in medical research projects.

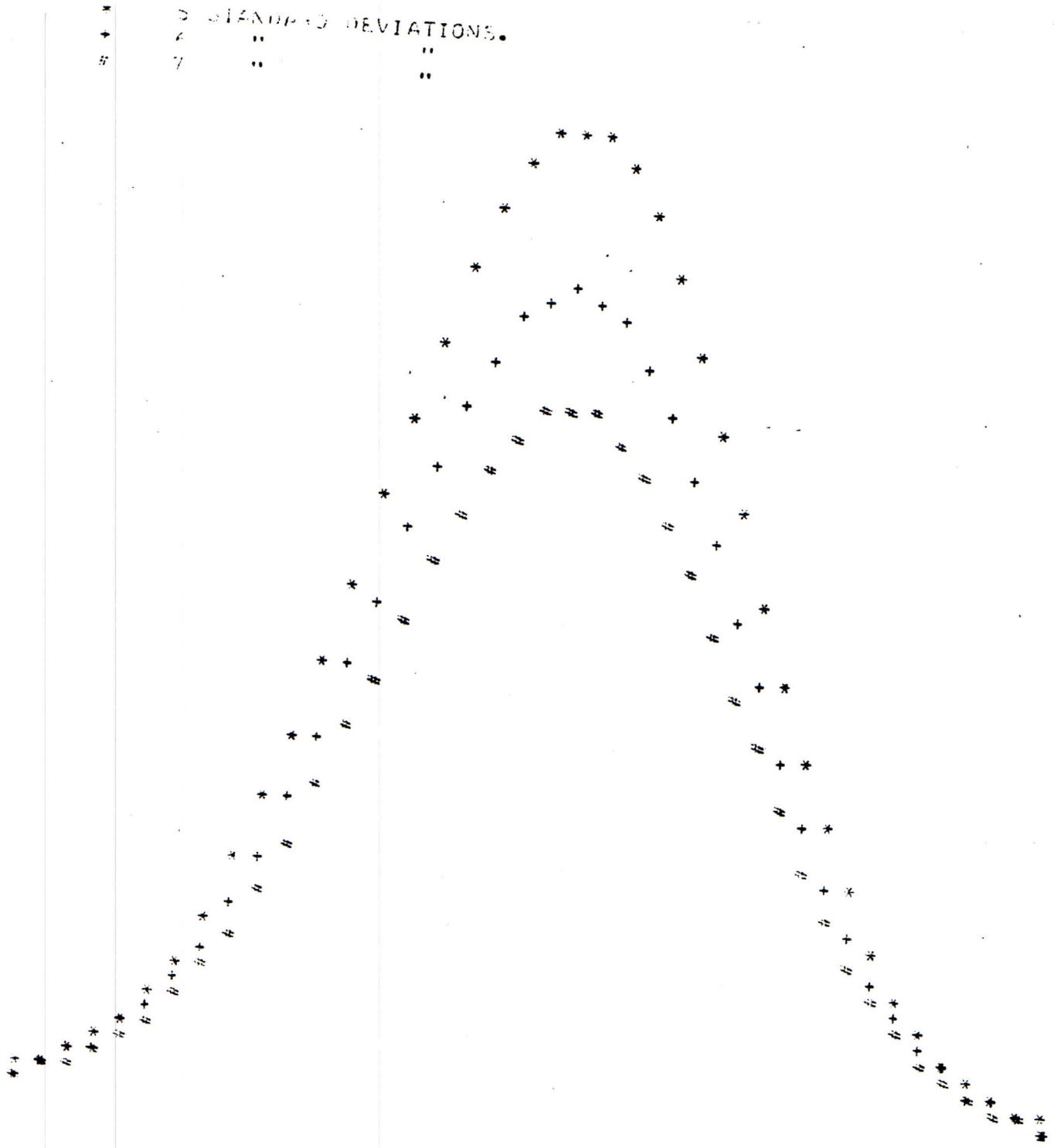
The package includes:

1. Plotting the Normal Curve for instruction purposes
2. Calculation of the mean and standard deviation values for a single sample.
3. Student's 't' Analysis.
4. 2 x 2 Chi-squared analysis together with an open ended chi-squared programme for testing goodness of fit.
5. Least squares correlation programme together with a Spearman-rho correlation by rank.
6. Analysis if Variance for two samples with one criterion of classification

Although designed for medical purposes there is no reason why these routines cannot be used in other faculties

Normal Curve Plot Programme One

The graph plot facilities offered by FOCAL have been utilized to produce a pedagogic programme to demonstrate the properties of a Normal Curve.


STANDARD DEVIATIONS.
"
"
"

Programme ONE

```
C-FOCAL., 1968

01.90S X=FSQT(2*3.142)

02.10A "STANDARD DEVIATION VALUE",SD
02.20F T=-2,,1,2;T "*",!;F Y=0,[800/(SD*X)]*FEXP<-T↑2>;T " "
*
```

The Normal Curve formula:

$$y = \frac{1}{\sigma\sqrt{2\Pi}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}$$

has been modified to:

$$y = \frac{800}{\sigma\sqrt{2\Pi}} e^{-t^2}$$

Where 800 represents the weighting factor to enable a reasonable curve size to be generated. The power of the exponential is condensed from $-(x-\bar{x})^2/2\sigma^2$ to $-t^2$ whose order of magnitude is represented by the FOR function spread of -2 to +2 in 0.1. increments.

Calculation of Mean ($\bar{x}$) and standard deviation (σ). Programme Two

Perhaps the most frequently required statistical parameters are the mean and standard deviation values for a single set of figures.

Using a linear array for storing individual items a sample size of over 100 figures can be accommodated. The mean is calculated as:

$$\bar{x} = \sum x / n$$

where 'n' is the number of items 'x' in the sample

The standard deviation is calculated as:

$$o = \sqrt{\sum x^2 / n - \bar{x}^2}$$

Student's 't' Analysis. Programme Three.

To assess the significance of the difference of the means of two

Programme TWO

```
*W
C-FOCAL., 1968

01.50C "ROUTINE FOR FINDING MEAN AND STANDARD DEVIATION"

02.10A "HOW MANY SAMPLES",N,!
02.20S Y=0;S Z=0;T "GIVE ",N,"VALUES",!
02.30F A=1,1,N;A X(A)
02.50F A=1,1,N;S Y=Y+X(A)
02.60S M=Y/N;T !, "                MEAN",M,!
02.70F A=1,1,N;S Z=Z+[X(A)]↑2
02.80S S=FSQT(Z/N-M↑2);T "STANDARD DEVIATION",S,!
*
```

Programme THREE

```
 W
C-FOCAL., 1968

01.50C "STUDENT'S T CALCULATION

02.10A "NUMBER IN GROUP ONE",N,!;T "GIVE",N," VALUES OF FIRST GROUP",!
02.30F  A=1,1,N;A X(A)
02.60T !, "GIVE",N," VALUES OF SECOND GROUP",!;S L=0;S Q=0;S R=0
02.70F  A=1,1,N;A Y(A)
02.80F  A=1,1,N;S Q=Q+X(A)

03.10F  A=1,1,N;S R=R+Y(A)
03.30F  A=1,1,N;S L=L+[X(A)-Y(A)]↑2
03.40S M=Q/N;S U=R/N;S S=FSQT([L/N]-<M-U>↑2)
03.50S T=<FABS(M-U)*FSQT[N-1]>/S;T !,"T VALUE",T," FOR",N-1," D.F.",!!
*

 GO
NUMBER IN GROUP ONE:10
GIVE=+   10.0000 VALUES OF FIRST GROUP
:7 :8 :6 :3 :9 :7 :9 :5 :7 :7
GIVE=+   10.0000 VALUES OF SECOND GROUP
:9 :8 :8 :4 :8 :7 :9 :7 :7 :6
T VALUE=+    1.3416 FOR=+    9.0000 D.F.

*
```

samples the following formula to calculate Student's 't' value is used:

$$t = \frac{(\bar{x} - m)^2 \sqrt{n}}{\sigma}$$

where 'm' is the mean value and $\bar{x}$ is the population mean value whose standard deviation is σ .

The two samples x and y are fed into the programme and stored in linear arrays. The means of the two samples, $\bar{x}$ and $\bar{y}$ are first calculated followed by the sum of the differences $\sum(x - y)^2$ The estimate of the population standard deviation is calculated as:

$$\sigma = \sqrt{\frac{\sum(x-y)^2}{n} - (\bar{y}-\bar{x})^2}$$

This is then used in the final equation for estimating 't'

Chi-squared Analysis. Programme Four

This programme calculates chi-square, and the associated two-tailed probability value, for the 2 x 2 contingency table where the number of cases are denoted by a, b, c and d as follows:

		X	Not X
	Y	a	b
Not	Y	c	d

Where N represents the total number of cases (a+b+c+d). The probabilities of the results arising with a null hypothesis (two tailed test) are given.

If it is felt necessary to include Yate's Correction in the formula, paragraph 2. 40 in the programme should be altered to read :

X = N (FABS [B*C-A*D] - N/2)↑2 / (a+b)*(a+c)*(c+d)*(b+d)

Programme FOUR

```
C-FOCAL., 1968

01.50C 2X2 CHI SQUARED GIVING SIGNIFICANCE LIMITS.

02.10T "GIVE ME THE SAMPLES IN THE FOLLOWING ORDER"
02.20T !!, "A  B",!,"C  D",!!
02.30A "A",A," B",B,!,"C",C," D",D;S N=A+B+C+D;T !,"GRAND TOTAL",N
02.40S X=[(B*C-A*D)↑2*N]/(A+B)*(C+D)*(A+C)*(B+D)
02.50IF (X-6.635) 3.1;
02.60T !,"CHI SQUARE",X," SIGNIFICANT AT 1% LEVEL";QUIT

03.10IF (X-3.841) 3.3;
03.20T !,"CHI SQUARE",X," SIGNIFICANT AT 5% LEVEL";QUIT
03.30T !,"CHI SQUARE NOT SIGNIFICANT";QUIT
*

GO
GIVE ME THE SAMPLES IN THE FOLLOWING ORDER

A  B
C  D

A:12   B:150
C:4   D:200
GRAND TOTAL=+ 366.0000
CHI SQUARE=+    6.4076 SIGNIFICANT AT 5% LEVEL*
```

Programme FIVE

```
W
C-FOCAL., 1968

02.10C CHI SQUARED PROGRAMME ALLOWING CONTINUATION.
02.20A " HOW MANY PREVIOUS SAMPLES HAVE YOU GIVEN",P,!
02.30A " WHAT IS THE CONTINUATION FIGURE ",C,!
02.40A " HOW MANY SAMPLES DO YOU WANT TO GIVE NOW",N,!
02.50T " GIVE ME",N," OBSERVED READINGS",!
02.60F A=1,1,N;A X(A)
02.80T !;T "O.K. NOW GIVE ME",N," EXPECTED READINGS",!
02.90F A=1,1,N;A Y(A)

03.20F A=1,1,N;S CI=CI+[X(A)-Y(A)]↑2/Y(A)
03.30T !,"" CHI SQUARED RESULT FOR ",(P+N)-1," D.F.",CI+C,!!
03.40S CI=0
*



GO
HOW MANY PREVIOUS SAMPLES HAVE YOU GIVEN:0
WHAT IS THE CONTINUATION FIGURE :0
HOW MANY SAMPLES DO YOU WANT TO GIVE NOW:10
GIVE ME=+   10.0000 OBSERVED READINGS
:200 :1150 :3246 :5100 :6300 :5200 :3000 :1300 :400 :120
O.K. NOW GIVE ME=+   10.0000 EXPECTED READINGS
:150 :1200 :3300 :5500 :6200 :5000 :2500 :1250 :450 :100
 CHI SQUARED RESULT FOR =+    9.0000 D.F.=+ 169.8930

*
```

A test for goodness of fit (Programme Five) uses the formula:

$$X^2 = \sum \frac{(o-e)^2}{e}$$

The programme is open ended and allows an indefinite number of observed (o) and expected (e) readings to be fed in Over 100 items per sample can be stored for computation with the facility, in the case of very large series, to split the sample and carry over the intermediate value for the Chi-squared figure.

Correlation

1 Least squares (Programme Six)

This method uses the formula:

$$R = \frac{\sum xy}{\sqrt{\Sigma x^2 \quad \Sigma y^2}}$$

2 Spearman - rho Correlation by rank, for two untied series (Programme Seven)

This uses the formula:

$$R = \frac{6 \sum D^2}{n^3 - n}$$

where D is the corresponding sample difference

The two sample series can first be ranked by using the ranking precursor programme (Programme Eight). In this programme the single sample of items is ranked by successive subtraction, noting in an auxillary linear array the order in which the items attain zero. Ties within individual samples are not tolerated.

Programme SIX

```
C-FOCAL., 1966

11.10T "PROGRAM FOR 'R'"
11.20A !, "HOW MANY SUBJECTS",N,!
11.30 T "GIVE",N," VALUES",!,"1ST.SAMPLE",!;S X=0;S Y=0;S M=0;S P=0;S Q=0
11.40F A=1,N;A X(A)
11.60F A=1,1,N;S X=X+X(A)
11.70T !, "2ND. SAMPLE",!
11.80F A=1,N;A Y(A)
11.90F A=1,N;S Y=Y+Y(A)

12.20F A=1,1,N;S M=M+X(A)*<Y(A)>
12.50F A=1,1,N;S P=P+(X(A))↑2
12.70F A=1,1,N;S Q=Q+(Y(A))↑2
12.80S P=FSQT(N*P-X↑2);S Q=FSQT(N*Q-Y↑2);S R=(N*M-X*Y)/P*Q;T !!, "R",R,!
*
```

```
GO
PROGRAM FOR 'R'
HOW MANY SUBJECTS:7
GIVE=+    7.0000 VALUES
1ST.SAMPLE
:-3 :-2 :-1 :0 :1 :2 :3
2ND. SAMPLE
:-4.2 :-3.1 :.6 :-.1 :1.5 :3.8 :3.1

R=+    0.9452
*
```

Programme SEVEN

```
C-FOCAL., 1968

01.80C SPEARMAN -RHO RANK CORRELATION.

02.10A "NUMBER OF SAMPLES TO BE USED ",N,!
02.20T " O.K. GIVE ME ",N, " FIRST VALUES",!;S D2=0;S R=0
02.30F A=1,1,N;A X(A)
02.50T !;T " O.K. NOW GIVE ME ",N, " SECOND VALUES",!
02.60F A=1,1,N;A Y(A)
02.80F A=1,1,N;S D2=D2+<Y[A]-X(A)>↑2
02.90S R=1-[(6*D2)/<N↑3-N>];T !!," RHO-COEFFICIENT ",R,!!
*


GO
NUMBER OF SAMPLES TO BE USED :10
O.K. GIVE ME =+  10.0000 FIRST VALUES
:2 :5 :6 :9 :8 :4 :3 :1 :10 :7
O.K. NOW GIVE ME =+  10.0000 SECOND VALUES
:3 :6 :5 :7 :8 :4 :2 :1 :9 :10

RHO-COEFFICIENT =+  0.8909

*
```

*W

Programme EIGHT

```
C-FOCAL., 1968

02.10A "HOW MANY SAMPLES FOR RANKING",N
02.20T % 8.04, !, "GIVE ME",N," VALUES"
02.30F A=1,N;A X(A)
02.40T !!;S P=0
02.50F A=1,N; DO 2.7
02.60T "LARGEST VALUE",P;S 0=N;S S=1; GOTO 4.11
02.70IF (P-X(A)) 2.8;RETURN
02.80S P=X(A);RETURN

04.11F A=1,N;DO 4.3
04.20S A=1;S S=S+1;G 4.11
04.30IF (X(A)-S) 4.11,4.4;RETURN
04.40S X(A)=P+1;S Y(A)=0;S 0=0-1;GOTO 5.1

05.10IF (0) 6.1,6.1;RETURN

06.10T !!,"RANK ORDER AS FOLLOWS",!;DO 7

07.10F A=1,N;T % 2, Y(A)
07.20 QUIT
*

 GO
HOW MANY SAMPLES FOR RANKING:10
GIVE ME=+  10.0000 VALUES:6 :4 :12 :8 :9 :2 :7 :5 :10 :3

LARGEST VALUE=+  12.0000

RANK ORDER AS FOLLOWS
=+ 6=+ 8=+ 1=+ 4=+ 3=+10=+ 5=+ 7=+ 2=+ 9*
```

Analysis of Varianee (Programme Nine.)

This programme is restricted to 2 samples x and y, where one criterion of classification is required. The samples are fed into two linear arrays n_1 and n_2. Unequal sample sizes are tolerated. The sample means and the standard deviation values for the samples X and y are printed out prior to construction of the analysis of variance table. The following formulae are used to compute:

1. The between sample sum of the squares:

$$SS = n\left(\left[\Sigma x/n - \Sigma x + \Sigma y/n_1 + n_2\right]^2 + \left[\Sigma y/n - \Sigma x + \Sigma y/n_1 + n_2\right]^2 \right)$$

In a two sample analysis of Variance where d. f. = 1 the between sample of variance is equal to SS.

2. The within sample sum of the squares:

$$WS = \sum(\bar{x} - x)^2 + \sum(\bar{y} - y)^2$$

- and the within sample variance = $$\frac{\sum(\bar{x} - x)^2 + \sum(\bar{y} - y)^2}{(n_1 - 1) + (n_2 - 1)}$$

The quotient of SS and WS gives the 'F' ratio figure for tabular comparison.

Programme NINE

```
C-FOCAL,, 1968

02.10C ANALYSIS OF VARIANCE: 2 SAMPLES: UNEQUAL SIZE
02.11S W1=0;S W2=0;S Y1=0;S X1=0;S D1=0;S D3=0
02.20A "HOW MANY ITEMS IN FIRST SAMPLE?",N1
02.30T ! "OK GIVE ME",N1," ITEMS",!!,"FIRST SAMPLE"
02.40F A=1,N1;A X(A)
02.50F A=1,N1;S X1=X1+X(A)
02.60S X2=X1/N1;T !!,"     FIRST SAMPLE MEAN:",X2
02.62F A=1,N1;S D1=D1+(X(A))↑2
02.64S D2=FSQT(D1/N1-X2↑2);T "     STANDARD DEVIATION:",D2,!!
02.70A "HOW MANY ITEMS IN SECOND SAMPLE",N2,!
02.80T "OK GIVE ME ",N2," ITEMS",!!,"SECOND SAMPLE"
02.90F A=1,N2;A Y(A)

03.10F A=1,N2;S Y1=Y1+Y(A)
03.20S Y2=Y1/N2;T !! "     SECOND SAMPLE MEAN:",Y2
03.21F A=1,N2;S D3=D3+(Y(A))↑2
03.22S D4=FSQT(D3/N2-Y2↑2);T "     STANDARD DEVIATION:",D4
03.30S T1=X1+Y1;S G4=T1/(N1+N2);T !! "     GRAND AVERAGE",GA,!!

04.10T "
04.20S S1=(X2-GA)↑2*N1;S S2=(Y2-GA)↑2*N2;S SS=S1+S2     SUMS OF SQUARES     D.F.     VARIANCE",!
04.30T "BETWEEN SAMPLES:     ",SS,"     ",1,"     ",SS

05.10F A=1,N1;S W1=W1+(X2-X(A))↑2
05.20F A=1,N2;S W2=W2+(Y2-Y(A))↑2
05.40S WS=W1+W2;T !! "WITHIN  SAMPLES:     ",WS,"     ",(N1+N2)-2,
05.50T "     ",WS/[(N1+N2)-2],!!!

06.10T  "
06.20T "TOTAL:     ",SS+WS,"     ",(N1+N2)-1,"     ",
06.30T !!"     F RATIO:     ",SS/<WS/([N1+N2]-2)>,!!!
*
```

```
GO
HOW MANY ITEMS IN FIRST SAMPLE?:6
OK GIVE ME=+    6.0000 ITEMS

FIRST SAMPLE:6 :12 :9 :8 :11 :9

    FIRST SAMPLE MEAN:=+    9.1667      STANDARD DEVIATION:=+    1.9508

HOW MANY ITEMS IN SECOND SAMPLE:6
OK GIVE ME =+    6.0000 ITEMS

SECOND SAMPLE:9 :4 :7 :11 :5 :10

    SECOND SAMPLE MEAN:=+    7.6667     STANDARD DEVIATION:=+    2.5604

          GRAND AVERAGE=+    8.4167

                    SUMS OF SQUARES          D.F.        VARIANCE

BETWEEN SAMPLES:      =+    6.7500      =+    1.0000    =+    6.7500

WITHIN  SAMPLES:      =+   62.1666      =+   10.0000    =+    6.2167

          TOTAL:      =+   68.9166      =+   11.0000    =+    6.2651


      F RATIO:    =+    1.0858
```