



# DECUS

## PROGRAM LIBRARY

DECUS NO.	8-321
TITLE	REG-2 (CURVILINEAR REGRESSION) REG-4 (LINEAR REGRESSION)
AUTHOR	Submitted by: A.J.P. Gore
COMPANY	The Nature Conservancy Grange-over-Sands Lancashire, England
DATE	November 27, 1969
SOURCE LANGUAGE	FORTRAN D

### ATTENTION

This is a USER program. Other than requiring that it conform to submittal and review standards, no quality control has been imposed upon this program by DECUS.

The DECUS Program Library is a clearing house only; it does not generate or test programs. No warranty, express or implied, is made by the contributor, Digital Equipment Computer Users Society or Digital Equipment Corporation as to the accuracy or functioning of the program or related material, and no responsibility is assumed by these parties in connection therewith.

# MEMORANDUM

FOR THE RECORD



REG-2 (CURVILINEAR REGRESSION) REG-4 (LINEAR REGRESSION)

DECUS Program Library Write-up

DECUS No. 8-321

REG-2 - CURVILINEAR REGRESSION FOR ANY NUMBER OF DATA

SUMMARY

This program computes the values  $a$ ,  $b$  and  $c$  for a second degree polynomial equation of the form:

$$Y = a + bX + cX^2$$

where  $X$  and  $Y$  can represent any number of values. The sum of squares of  $Y$ , and the sum of the products of  $X^2Y$  and  $XY$  are also printed out, which are required to determine the sum of squares due to regression. Positive or negative values of  $X$  and  $Y$  may be used.

TAPES REQUIRED

Form of program tape - The program is written in the PDP-8 FORTRAN-D language, and is in the source language.

Form of data tape - The data to be analyzed should be punched onto paper tape, the  $X$  values first, and then the  $Y$  values. Both should be expressed as variables, not integers. At the beginning of the program the total number of  $X$  or  $Y$  variables should be punched, in the form of an integer (i.e., if there are 40 pairs of  $X$  and  $Y$ , punch 40). This value represents  $n$  in the program tape and enables the program to run for any number of variables.

OPERATING INSTRUCTIONS

```
.FORT          Program tape in high-speed reader
*OUT-S:REG 2
*
*IN-R:
* ↑
↑
*READY        Data tape in high-speed reader
↑
```

If the program tape has already been compiled onto the disk, it may be called back into core as follows:

```
.FOSL
*IN-S:REG 2
*OPT-
* ↑
*READY
↑
```

## OUTPUT

The program prints out the values under:

A, B, C, PXZY, PXY and PYZ

## STORAGE AND LIMITATION

X and Y values can be stored with up to 3 digits after the decimal point without losing too much precision (error in the range of  $10^4$ ) for fairly low numbers of variables, though it may be slightly higher for high numbers).

## METHOD

See "Principles and Procedures of Statistics", p. 338, Steel, R. G. D. and Torrie, J. H.

L

C CURVILINEAR REGRESSION 2ND METHOD

DIMENSION X1(25), Y(25)

DO 2Ø I=1, 25

READ 2, 1Ø1, X1(I)

1Ø1

FORMAT(/, E)

2Ø

CONTINUE

DO 3Ø J=1, 25

READ 2, 1Ø2, Y(J)

1Ø2

FORMAT(/, E)

3Ø

CONTINUE

READ 2, 1Ø4, K

1Ø4

FORMAT(/, I)

SY=Ø.Ø

SX1=Ø.Ø

SX2=Ø.Ø

SX1Y=Ø.Ø

SY2=Ø.Ø

SX3=Ø.Ø

SX4=Ø.Ø

SX2Y=Ø.Ø

DO 1Ø N=1, 25

SY=SY+Y(N)

SX1=SX1+X1(N)

SX1Y=SX1Y+X1(N)\*Y(N)

IF(X1(N) )7Ø, 71, 71

71

SX2=SX2+X1(N)\*\*2

SX3=SX3+X1(N)\*\*3

SX4=SX4+X1(N)\*\*4

SX2Y=SX2Y+(X1(N)\*\*2)\*Y(N)

GO TO 1ØØ

7Ø

X1(N)=-X1(N)

SX2=SX2+X1(N)\*\*2

SX3=SX3-X1(N)\*\*3

SX4=SX4+X1(N)\*\*4

SX2Y=SX2Y+(X1(N)\*\*2)\*Y(N)

1ØØ

IF(Y(N) )8Ø, 81, 81

8Ø

Y(N)=-Y(N)

81

SY2=SY2+Y(N)\*\*2

1Ø

CONTINUE

X1M=SX1/25.Ø

X2M=SX2/25.Ø

YM=SY/25.Ø

PX3=SX3-SX1\*SX2/25.Ø

PX2Y=SX2Y-SX2\*SY/25.Ø

PXY=SX1Y-X1M\*YM\*25.Ø

IF(X1M)9Ø, 91, 91

```

90      X1MP=-X1M
        PX2= SX2-(X1MP**2)*25.0
        GO TO 200
91      PX2= SX2-(X1M**2)*25.0
200     IF(SX2)110,111,111
110     SX2P=-SX2
        PX4= SX4-(SX2P**2)/25.0
        GO TO 300
111     PX4= SX4-(SX2**2)/25.0
300     IF(PX3)120,121,121
120     PX3P=-PX3
        D=(PX2*PX4)-PX3P**2
        GO TO 400
121     D=(PX2*PX4)-(PX3**2)
400     PA=(PXY*PX4-PX2Y*PX3)/D
        PB=(PX2*PX2Y-PX3*PXY)/(D*10.0**K)
        C=PB
        B=PA
        A=YM-(C*X2M*10.0**K)-(B*X1M)
        A=A*10.0**K
        TYPE 103,A,B,C
103     FORMAT(/,"A",E,"B",E,"C",E,)
        TYPE 105,PX2Y,SY2,PXY
105     FORMAT(/,"PX2Y",E,"SY2",E,"PXY",E)
        END

```

\*

# REG-4 - LINEAR REGRESSION FOR SIMPLE LINEAR, EXPONENTIAL AND HYPERBOLIC RELATIONSHIPS

## SUMMARY

This program computes the values of:

- |     |   |                                    |
|-----|---|------------------------------------|
| (a) | a and b in the equation                   | $Y = a + b X$                      |
| (b) | $\log_e a$ and $\log_e b$ in the equation | $\log_e Y = \log_e a + X \log_e b$ |
| (c) | $\log_e a$ and b in the equation          | $\log_e Y = \log_e a + b \log_e X$ |

Any number of pairs of X, the independent variable, and Y, the dependent variable, may be used and they may be either positive or negative. Any number of regressions may be calculated consecutively, using any or all of the above relationships.

## TAPES REQUIRED

Form of program tape - The program is written in the PDP-8 FORTRAN-D language, and is in the source language.

Form of data tape - The data tape contains the X, Y values for the first regression calculation in the form:

$$X_1 Y_1 \quad X_2 Y_2 \quad \dots \quad X_n Y_n$$

If more than one regression calculation is required, subsequent sets of X and Y follow.

## OPERATING INSTRUCTIONS

.FORT	Program tape in high-speed reader
*OUT-S:REG 4	
*	
*IN-R:	
* ↑	
*READY	Data tape in high-speed reader
↑	

The program will then pause for entry, on the keyboard, of the sequence:

p q n "Return"

where p is the number of regressions to be calculated;

q is 0 for linear regression with no transformation

1st regression      1 for  $\log_e$  transformation of Y (exponential)

                         2 for  $\log_e$  transformation of X and Y (hyperbolic)

n is number of pairs of X, Y variables

if p = 1 the program will pause for entry of the relevant q n at the beginning of each regression calculation

#### OUTPUT

For each regression the program prints the form of the relationship,  $\bar{X}$ ,  $\bar{Y}$ ,  $SSX$ ,  $SSY$ ,  $SXY$ ,  $a$ ,  $b$ ,  $r$ ,  $r^2$ , regression SS, residual SS, mean square deviation, and number of pairs of X, Y variables.

Where transformations have been carried out, the appropriate transformed parameters will be output.

#### METHOD

Snedecor "Statistical Methods".

```

L
C SYKES REGRESSION Y ON X WITH LOG TRANSFORM OPTIONS REG4
  N5=0
100 FORMAT (/, E, E)
101 FORMAT (I, I)
  ACCEPT 101, N1
128 XBAR=0.
  SSX=0.
  YBAR=0.
  SSY=0.
  SSXY=0.
  ACCEPT 101, N2, N3
  DO 1 J=1, N3
  E=J
  READ 2, 100, X, Y
  IF (N2-1) 110, 105, 107
105 Y=LOGF(Y)
  GO TO 110
107 Y=LOGF(Y)
  X=LOGF(X)
110 SSX=SSX+(X-XBAR)*(X-XBAR)*(1.-1./EJ)
  SSY=SSY+(Y-YBAR)*(Y-YBAR)*(1.-1./EJ)
  SSXY=SSXY+(X-XBAR)*(Y-YBAR)*(1.-1./EJ)
  XBAR=XBAR+(X-XBAR)/EJ
  YBAR=YBAR+(Y-YBAR)/EJ
 1 CONTINUE
  IF (N2-1) 129, 132, 133
129 FORMAT (/, /, "Y=A+B*X", E)
  TYPE 129
  GO TO 134
132 FORMAT (/, /, "LN(Y)=LN(A)+LN(B)*X", E)
  TYPE 132
  GO TO 134
133 FORMAT (/, /, "LN(Y)=LN(A)+B*LN(X)", E)
  TYPE 133
134 B=SSXY/SSX
  A=YBAR-B*XBAR
  R=SSXY/SQTF(SSX*SSY)
  RSQ=R*R
  SST0=SSXY*SSXY/SSX
  SSFR=SSY-SST0
  SDEV=SSFR/(EJ-2.)
116 FORMAT (/, "XBAR", E)
117 FORMAT (/, "SUM SQUARES X", E)
118 FORMAT (/, "SUM SQUARES Y", E)
119 FORMAT (/, "YBAR", E)
141 FORMAT (/, "SUM CROSS PROD", E)

```

```
120 FORMAT (/, "A", E)
121 FORMAT (/, "B", E)
122 FORMAT (/, "R", E)
123 FORMAT (/, "RSQ", E)
124 FORMAT (/, "SS TO REG", E)
125 FORMAT (/, "SS FROM REG", E)
126 FORMAT (/, "MEAN SQ DEV", E)
127 FORMAT (/, "NO PAIRS OBS", E)
TYPE 116, XBAR
TYPE 117, SSX
TYPE 118, SSY
TYPE 119, YBAR
TYPE 141, SSXY
TYPE 120, A
TYPE 121, B
TYPE 122, R
TYPE 123, RSQ
TYPE 124, SSTO
TYPE 125, SSFR
TYPE 126, SDEV
TYPE 127, EJ
N5=N5+1
IF (N5-N1) 128, 2, 2
2 CONTINUE
END
```

\*P