DECUS
PROGRAM LIBRARY

DECUS NO.	FOCAL8-63
TITLE	CURFIT
AUTHOR	Donald L. Shirer
COMPANY	Valparaiso University Valparaiso, Indiana
DATE	October 22, 1969
SOURCE LANGUAGE	FOCAL

CURFIT

DECUS Program Library Write-up

DECUS No. FOCAL8-63

Abstract: CURFIT is a program written in the FOCAL language which fits weighted or unweighted data to a straight line on a Cartesian, log-log or semilog graph. It calculates the slope and intercept of the line, the standard error in these values, plus other measures of the "goodness" of fit. Values may be added or deleted from the data list easily, and there is no limit to the number of sample data pairs.

Operating Procedure:

1) Load the FOCAL system, keeping the extended functions.

2) Type ERASE ALL, then load the CURFIT source tape using either reader.

   a) High-speed reader: Place tape in reader, type * and RETURN.

   b) Teletype reader: Place tape in reader, flip reader switch to START. The program tape supplied has inserted blanks which prevent the reader from getting ahead of the printer. If you use a copy without the blanks, occasionally flip switch OFF (every 10 lines or so) wait a second, then turn to START again. When tape has fed through, turn switch to FREE and remove tape.

3) Type GO and RETURN.

4) Program will ask for your choice of options. Type +1 for NEW data, $\emptyset$ for OLD data modification, then SPACE. If you typed $\emptyset$, go to step 7.

5) Program will ask for type of fit. Type $\emptyset$ for LINEAR, +1 for LOG-LOG, -1 for SEMILOG, then SPACE.

6) Program will ask for weight option. Type $\emptyset$ if you wish to enter the weight w_i for each data pair, +1 if you wish to choose the EQUAL WT option and set all weights equal to unity, or -1 if you desire the COUNT WT option for counter data with a Poisson distribution.

7) Enter the data pairs x_i, y_i to be added to the data list as the program asks for them. SPACE after each entry. If the WEIGHTED option has been chosen in step 6, you will be asked to enter w_i after you have entered x_i and y_i. You may enter the numbers with or without a decimal point, or in exponential power-of-ten notation. (Thus 100 ; 100.0 ; 1E2 ; 1.0E+2 are all valid and equivalent data entries.)

8) If you make a typing error and discover it <u>before</u> you enter the number by spacing, you may erase it by typing ← immediately, then space and re-enter the correct value.

9) You may delete a value from the data list (perhaps because you later discovered that it was in error) by one of the following procedures:

   a) WEIGHTED OPTION. If you discover an error <u>after</u> you have entered x_i and y_i, but <u>before</u> you enter w_i, enter $\emptyset$ for w_i and that line will be ignored. Re-enter correct value on next line.

b) WEIGHTED OPTION. To delete a previous line, retype the x_i and y_i exactly as you entered them before (including the error, if any) but enter as w_i the negative of the previous weight. This line will exactly cancel the effect of the previous line, and the value of N will be adjusted accordingly.

c) COUNT WT OPTION. To delete a previous line, retype it with the same x_i but the negative value of the count y_i.

d) EQUAL WT OPTION. You are stuck. The easiest thing to do is start over by pressing CTRL-C and then typing GO again. If you do not want to waste the time you have spent entering data, a procedure is given in Appendix C which will enable you to recover your error and delete a previously entered data pair.

10) When all data pairs have been entered, type TOTAL when asked for the next x_i, then SPACE.

11) The parameters and criteria of fit will be typed out immediately. See the following pages for their definitions and interpretation.

12) Samples of the printout for several option choices are shown in Appendix A.

13) After the printout, the program begins again at step 4. If the OLD RUN option is now chosen, you may add or delete data pairs from the previous data list. An example is given in Appendix B. (Since it is easier to delete data pairs if the WEIGHTED option is chosen, it is suggested that you select this option if you intend to modify the data list on a second run.)

14) To escape from the program and return to FOCAL command mode, press the CTRL-C keys simultaneously.

Weighted Fits:

CURFIT will evaluate the slope B and intercept A of the straight line which is the "best" fit (by the criteria of least squares) to N data points x_i, y_i, each with a weighting factor w_i. The following well-known formulas are used, which assume that the uncertainty in y_i is much larger than that in x_i:

$$\text{if } DX = \Sigma W_i \Sigma W_i X_i^2 - (\Sigma W_i X_i)^2 \quad \text{then}$$

$$A = (\Sigma W_i Y_i \Sigma W_i X_i^2 - \Sigma W_i X_i \Sigma W_i X_i Y_i)/DX \tag{1}$$

$$B = (\Sigma W_i \Sigma W_i X_i Y_i - \Sigma W_i X_i \Sigma W_i Y_i)/DX \tag{2}$$

The standard error in the coefficients is given by

$$SA = S\,(\Sigma W_i X_i^2/DX)^{\frac{1}{2}} \qquad SB = S\,(\Sigma W_i/DX)^{\frac{1}{2}} \tag{3}$$

where

$$S = (\Sigma W_i\,(Y_i - A - B X_i)^2/(N-2))^{\frac{1}{2}} \tag{4}$$

is the best estimate of the standard deviation of the data points from the straight line. The desired relation between X and Y is thus

$$Y = (A \pm SA) + (B \pm SB) \cdot X \qquad \text{(Linear Fit)} \tag{5}$$

To sketch the straight line, two points on the line are needed. These can be the points (0,A) and $(\bar{X},\bar{Y})$, where

$$\bar{X} = \Sigma W_i X_i / \Sigma W_i \qquad \bar{Y} = \Sigma W_i Y_i / \Sigma W_i \tag{6}$$

Criteria of Fit:

The correlation coefficient $R = \dfrac{\Sigma W_i (X_i-\bar{X})(Y_i-\bar{Y})}{(\Sigma W_i (X_i-\bar{X})^2 \Sigma W_i (Y_i-\bar{Y})^2)^{\frac{1}{2}}}$ (7)

is in many ways the best indicator of a linear relation between two variables. It is zero for no correlation and unity for perfect correlation.

If there is any functional relationship between Y and X, the slope of the line ought to differ from zero. The T-statistic for the Student Test of a significant departure from zero slope is given by

$$T = B/SB \tag{8}$$

for N-2 degrees of freedom. (When SB is zero, the fit is perfect, and

neither R nor T is calculated by CURFIT to avoid possible error messages.)

Forced Fit Through Origin:

Occasionally it is known with certainty that for X=0, Y=0. In addition to the "best fit" line, CURFIT calculates the slope of a line passing through the origin which best fits the given data, the standard error in this slope, and the corresponding T-statistic for a significantly non-zero slope (with N-1 degrees of freedom) from the formulas:

$$B_o = \Sigma W_i X_i Y_i / \Sigma W_i X_i^2 \; ; \quad SB_o = \frac{(\Sigma W_i Y_i^2 - B_o^2 \Sigma W_i X_i^2)^{\frac{1}{2}}}{(N-1)^{\frac{1}{2}} \; \Sigma W_i X_i^2} \; ; \quad T_o = B_o / SB_o \qquad (9)$$

If this feature will never be used, it can be suppressed by changing line 5.3 in the FOCAL program list to

```
5.30  GOTO 1.1
```

Weighting Factors:

If all data samples are of equal significance, then they should be weighted equally. CURFIT assigns a weight of +1 to each data pair if the user chooses the EQUAL WT option, so the user does not have to enter the weight himself. When some data samples are drawn from a more reliable distribution, they should have higher weight. The least-square formulas used assume that the weights are inversely proportional to the variance in the distribution from which the samples y_i are drawn, that is,

$$w_i = \text{constant}/S_{y_i}^2 \qquad (10)$$

where S_{y_i} is the standard deviation in all similar y_i samples. In practice, usually one can make only an estimate of the proper w_i, but this can significantly improve the values of A, B, SA and SB. The constant in (10) may be chosen to be any convenient value, perhaps +1.

Log Plots:

Semilog and log-log graphs are made by plotting not the original raw data x_i, y_i, but the converted data pairs X_i, Y_i obtained as follows:

Graph	$X_i =$	$Y_i =$	Equation corresponding to $Y = (A \pm SA) + (B \pm SB) \cdot X$
Semilog	x_i	$\ln(y_i)$	$y = e^{(A \pm SA)} \cdot e^{(B \pm SB) \cdot x}$ (exponential)
Log-Log	$\ln(x_i)$	$\ln(y_i)$	$y = e^{(A \pm SA)} \cdot x^{(B \pm SB)}$ (power law)

Whenever a log-conversion is made, the value of e^A is automatically included in the output.

If the weights were allowed to remain the same, the deviations of $\ln(y_i)$ from the straight line on a log plot will be minimized during the least squares fit. This is what most experimenters might do when making an "eyeball" fit to a semilog or log-log graph, but is not correct of the weights w_i of the raw data reflect the proper estimated variance in y_i as in (10). The variance of the converted data is

$$S_{Y_i} = S_{y_i} (dY/dy) = S_{y_i} (1/y_i) \tag{11}$$

so that the converted weights should be

$$W_i = W_{Y_i} = 1/S^2_{Y_i} = y_i^2/S^2_{y_i} = w_{y_i} y_i^2 \tag{12}$$

This conversion is made automatically by CURFIT whenever a SEMILOG or LOG-LOG option is specified. The higher weight for larger y_i just reflects the fact that the log scale is "squeezed together" at higher values of y so the fitted line must lie nearer the data points with large y.

Counting Statistics:

Data from particle counters and other samples with a Poisson distribution have a variance proportional to the number of counts y_i, so

$$S_{y_i} = \text{constant } (y_i)^{\frac{1}{2}} \tag{13}$$

and the proper weighting factor will be

$$w_i = \text{constant}/y_i \quad \text{(or just } 1/y_i) \tag{14}$$

This value will be selected automatically if the COUNT option is chosen, so the CURFIT user does not have to calculate or enter the weights himself. For example, when finding the decay curve of a radioactive isotope, the user would select the SEMILOG and COUNT options, enter the time after the

start of the experiment as x_i and the count accumulated during a constant interval starting at that time as y_i. Then the mean decay life of that isotope would be 1/B, where B is the slope of the exponential curve fitted to the data.

Modifying the Data List:

If after calculating the best fit to a set of data pairs, you wish to see if the fit improves when certain data is added or subtracted to the data list, you may choose the OLD RUN option when the program is ready for new instructions. New data points may be added to the previous list by simply typing them in when new data entry is requested. Data points may also be deleted from the previous list by following the procedure under ERROR CORRECTION in the operating instructions.
It is difficult to delete data if the EQUAL WT option has been chosen, so if you anticipate that this will be necessary, you may choose the weighted option and enter 1 for each weight during normal data entry.

Number of Data Points:

There is no practical maximum limit on the number of data pairs you may enter. However, you must enter at least three data pairs for the least squares formulas to be valid.

```
*C    APPENDIX A:  SAMPLES OF CURFIT OUTPUT
*G


ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN: 1
LINEAR(0), LOG-LOG(1), OR SEMILOG(-1) FIT: 0
WEIGHTED(0), EQUAL-WT(1), OR COUNTER(-1) DATA: 1

ENTER DATA, THEN TYPE 'TOTAL'
X:1       Y:5
X:2       Y:8
X:3       Y:11
X:TOTAL

N=     3
A = 0.200000E+01 +- SA= 0.000000E+00
B = 0.300000E+01 +- SB= 0.000000E+00
X = 0.200000E+01      Y = 0.800000E+01
R = 0.100000E+01      S = 0.000000E+00

FORCED THROUGH ORIGIN:
B = 0.385714E+01 +- SB= 0.247441E+00
B/SB= 0.155881E+02


ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN: 1
LINEAR(0), LOG-LOG(1), OR SEMILOG(-1) FIT: 1
WEIGHTED(0), EQUAL-WT(1), OR COUNTER(-1) DATA: 0

ENTER DATA, THEN TYPE 'TOTAL'
X:1       Y:4       W:2
X:2       Y:16      W:2
X:3       Y:36      W:1
X:TOTAL

N=     3
A = 0.138630E+01 +- SA= 0.321066E-02
FEXP(A)= 0.400002E+01
B = 0.199999E+01 +- SB= 0.323698E-02
X = 0.966681E+00      Y = 0.331966E+01
B/SB= 0.617857E+03
R = 0.100000E+01      S = 0.308397E-01

FORCED THROUGH ORIGIN:
B = 0.336217E+01 +- SB= 0.221305E+00
B/SB= 0.151925E+02


ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN:?01.00 @ 01.20
*
```

Formula of Fitted Curve

$y = 2.0 + 3.0\,x$

$y = 4.0\,x^{2.0}$

```
*C    APPENDIX B:    SAMPLE OF DATA MODIFICATION
*G


ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN: 1
LINEAR(0), LOG-LOG(1), OR SEMILOG(-1) FIT: -1
WEIGHTED(0), EQUAL-WT(1), OR COUNTER(-1) DATA: -1

ENTER DATA, THEN TYPE 'TOTAL'
X:1       Y:1000
X:2       Y:500
X:3       Y:240
X:4       Y:130
X:TOTAL

N=     4
A = 0.759931E+01 +- SA= 0.259878E-01
FEXP(A)= 0.199681E+04
B =-0.693678E+00 +- SB= 0.132075E-01
X = 0.173262E+01     Y = 0.639743E+01
B/SB=-0.525215E+02
R =-0.999638E+00     S = 0.532626E+00

FORCED THROUGH ORIGIN:
B = 0.270712E+01 +- SB= 0.105682E+01
B/SB= 0.256156E+01
```

y = 1997 exp(-.6937 x)

```
ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN: 0

ENTER DATA, THEN TYPE 'TOTAL'
X:1       Y:-1000
X:1       Y:1100
X:TOTAL

N=     4
A = 0.772648E+01 +- SA= 0.499077E-01
FEXP(A)= 0.226760E+04
B =-0.737233E+00 +- SB= 0.258555E-01
X = 0.169543E+01     Y = 0.647655E+01
B/SB=-0.285136E+02
R =-0.998772E+00     S = 0.105890E+01

FORCED THROUGH ORIGIN:
B = 0.277863E+01 +- SB= 0.110478E+01
B/SB= 0.251510E+01
```

Old Run option selected
Other options automatically chosen same as on previous run.
Remove original data pair
Substitute correct pair

y = 2268 exp(-.7372 x)

```
ENTER CURFIT OPTIONS
NEW(1) OR OLD(0) RUN:
```

APPENDIX C : To Delete a Data Pair When Using EQUAL WT Option.

The strategy is to change to the WEIGHTED option, then delete the data pair as in step 9(b) of the operating instructions.

1) When you discover that you need to delete a previous x_i, y_i, press CTRL-C to enter command mode.
2) Type the following line: SET WS = $\emptyset$; GOTO 2.2 Press RETURN.
3) Re-enter the incorrect line to be deleted. When w_i is requested, type -1.
4) Continue entering data. Since you are now using the WEIGHTED option, enter +1 for the weight of each data pair to be added to the data list.

APPENDIX D : Program Variables.

A New or Old Run switch. Later used as value of Intercept of best-fit line.

B Slope of best-fit line.

DX $\Sigma W_i \Sigma W_i X_i^2 - (\Sigma W_i X_i)^2$ as used in equations (1) (2) (3)

DY $\Sigma W_i \Sigma W_i Y_i^2 - (\Sigma W_i Y_i)^2$ used to simplify calculation of S.

N Number of data pairs entered - number of pairs deleted.

SA Standard error in A

SB Standard error in B

SW ΣW_i

TS Type of fit switch. $\emptyset$ for linear fit, +1 for log-log, -1 for semilog.

W Raw weight w_i. If TS $\neq \emptyset$, it is converted to W_i as in equation (12).

WS Weight option switch. $\emptyset$ or Weighted option, +1 for Equal Wt., -1 for Count Wt.

X, Y Originally the raw data x_i, y_i. If TS $\neq \emptyset$, they are converted to X_i, Y_i as shown on page 4. Later used for average values $\bar{X}$, $\bar{Y}$ in printout.

XX $\Sigma W_i X_i^2$

XY $\Sigma W_i X_i Y_i$

YY $\Sigma W_i Y_i^2$

SX $\Sigma W_i X_i$

SY $\Sigma W_i Y_i$

B/SB T statistic for Student Test
R Correlation coefficient
} These only appear in output.

S Standard deviation of data pairs from fitted line. See equation(4).

```
*C    APPENDIX E :  PROGRAM LISTING
*
*WRITE ALL
C-FOCAL,1969

01.10 T !!!!"ENTER CURFIT OPTIONS
01.20 A !"NEW(1) OR OLD(0) RUN",A
01.30 I (A)2.1,2.1
01.40 A !"LINEAR(0), LOG-LOG(1), OR SEMILOG(-1) FIT",TS
01.50 A !"WEIGHTED(0), EQUAL-WT(1), OR COUNTER(-1) DATA",WS

02.05 S N=0;S SW=0;S SX=0;S SY=0;S XX=0;S YY=0;S XY=0
02.10 T !!"ENTER DATA, THEN TYPE 'TOTAL'
02.20 S W=1;A !,?X?;I (X-0TOTAL)2.3,4.1
02.30 A ?     Y?;I (-WS)2.7,2.6
02.40 I (Y)2.5,3.9
02.50 S W=1/Y;S Y=FABS(Y);G 2.7
02.60 A ?     W?
02.70 I (W)3.1,2.2;S NN=N+1;G 3.3

03.10 S NN=N-1
03.30 I (TS)3.5,3.7,3.4
03.40 I (X)3.9,3.9;S X=FLOG(X)
03.50 I (Y)3.9,3.9;S W=W*Y*Y;S Y=FLOG(Y)
03.70 S N=NN;S SW=SW+W;S SX=SX+W*X;S SY=SY+W*Y
03.80 S XX=XX+W*X*X;S YY=YY+W*Y*Y;S XY=XY+W*X*Y;G 2.2
03.90 T "VALUE MUST BE +";G 2.2

04.10 T %4,!!,?N?
04.20 S X=SX/SW;S Y=SY/SW;S DX=SW*XX-SX↑2;S DY=SW*YY-SY↑2
04.30 S B=(SW*XY-SX*SY)/DX;S A=Y-B*X
04.40 S S=FSQT(FABS((DY-B*B*DX)/((N-2)*SW)))
04.50 S SA=S*FSQT(XX/DX);S SB=S*FSQT(SW/DX);T %,!,?A ?," +- ",?SA?
04.60 I (TS)4.7,4.8
04.70 T !,?FEXP(A)?
04.80 T !,?B ?," +- ",?SB?
04.90 T !,?X ?,?     Y ?

05.10 D 6;T !"R ",B*FSQT(DX/DY),?     S ?,!!"FORCED THROUGH ORIGIN:"
05.20 S B =XY/XX;S SB=FSQT((YY-B*B*XX)/((N-1)*XX))
05.30 D 4.8;D 6;G 1.1

06.10 I (SB)6.2,6.2;T !,?B/SB?
06.20 R
*
*
*C     GROUP 1 REQUESTS CHOICE OF OPTIONS
*C     GROUP 2 IS DATA ENTRY AND WEIGHT SETTING ROUTINE
*C     GROUP 3 CONVERTS DATA FOR LOG OPTIONS AND INCREMENTS SUMS
*C     GROUP 4 CALCULATES THE PARAMETERS AND CRITERIA OF FIT
*C     GROUP 5 CALCULATES THE PARAMETERS OF LINE FORCED THRU ORIGIN
*C     GROUP 6 IS SUBROUTINE USED TO PRINTOUT T STATISTIC
*
```