Documentation for PC-DATA™

Copyright

License

The original, and any copies of the program shall at all times be the sole property of Structural Reliability Consultants. Being a beta tester you agree after 12/31/90 to either erase (destroy) or return all existing copies of the beta test version 2.00β. If you want a copy of PC-DATA™ Version 2.00β to beta test contact:

Joseph Murphy, Ph.D., P.E
P.O. Box 56164
Madison, WI 53705
Phone 608-274-0527

Personal Computer Requirements

IBM® personal computer or compatible with CGA or EGA adapter.
PC-DATA™ requires an x87 math coprocessing chip!
256K memory, 1 Disk Drive

_____

Probability Distributions

The program gives graphical and maximum likelihood parameter estimation of three general probability distributions (with or without suspended data):

1. Two parameter Normal (2)*
2. Three parameter Lognormal (3)
3. Three parameter Weibull

NOTE: The three parameter Weibull cumulative distribution function (CDF) is defined:

$$F(x) = 1 - \exp\left[-\{(x-x_o)/(\eta-x_o)\}^{\alpha}\right]$$

Defined this way the scale parameter η (eta) is the .63 fractile of the estimated distribution!

The .05 fractiles of the distributions are:

$x_{.05} = \mu_x - 1.645\,\sigma_x$	Normal
$x_{.05} = x_o + \exp\left[\,\mu_{\ln(x-x_o)} - 1.645\,\sigma_{\ln(x-x_o)}\right]$	Lognormal
$x_{.05} = x_o + (\eta-x_o)\,(0.0513)^{1/\alpha}$	Weibull

Data Files

The data files are ASCII files. Data files with more than 140 points (but less than 1500) have to be inputted with a word processor. One data per line. PC-DATA™ will sort the data. Just make sure the extension file name is '.DAT'.

Ranking of Sample Data

Median Rank

$$P_j \cong (\,j - 0.3\,)/(\,n + 0.4\,)$$

The median ranking scheme has the advantage that half the data will be above and half the data below, and it has the least least squares of all ranking schemes (6).

---

* underlined numbers in parentheses refer to references.

Ranking of Sample Data containing Suspended Points

- Ordered Sample Values

  Completed (failed) Data : $x_1 \le x_2 \le ... \le x_n$

  Suspended Data : $y_1 \le y_2 \le ... \le y_m$

  where $N = n + m$

- Completed Data Order Number, j

  j = previous order number + increment

  where, when $x_i \le y_k \le x_{i+1}$

  $$\text{increment} = \frac{(N + 1) - \text{previous order number}}{1 + \text{Number of Sample Points with Value} > y_k}$$

Step Procedure (6)

1. previous order number PON=0, increment INC=1
   J is data point number, start with J=1

2. if data value is complete then PON=PON+INC and
   median rank is (PON - 0.3)/(N + 0.4)

3. if data value is suspended then there is no order number
   and no rank and INC= (N + 1 - PON)/(1 + N - J)

4. if J < N then J = J + 1, go to step 2

Example

J	Ordered Values	Order Number	Median Rank
1	x_1	1	.0673
2	x_2	2	.1635
3	y_1	-	-
4	x_3	3.125	.2716
5	y_2	-	-
6	y_3	-	-
7	x_4	4.7	.4231
8	x_5	6.275	.5745
9	y_4	-	-
10	y_5	-	-

Suspended Data

The parameter estimation accounts for 'suspended' data. Suspended data are inputted as negative values where the absolute value is the point at which the test was suspended (e.g. the load at which the member failed out of the test region, or the time the test was stopped, etc.).

Minimum Location Parameter x_o

To estimate parameters in 3 parameter distributions, the program gives the correlation coefficient of the least squares best fit straight line on 'probability plots' for 0,.1,.2, ... ,.9 times the lowest value (suspended or not). The user is to make an intelligent a priori decision of the location parameter, x_o. REMEMBER you are telling the world that there is ABSOLUTELY no value below x_o, EVER.

Graphical Confidence Bounds

Binomial-Type Bounds

One confidence bound technique uses the cumulative binomial distribution (1,4,5). This ranking procedure is general in nature and can be applied to any statistical distribution graphical analysis. For the one-sided confidence bound, the following equation is solved for R:

$$1 - C = \sum_{i=0}^{j-1} \binom{N}{i} R^i (1 - R)^{N-i} = (1 - R)^N \sum_{i=0}^{j-1} p_i$$

Where $\binom{N}{i} = \frac{N!}{i!\,(N-i)!}$ = binomial coefficients

C = one-sided confidence level, $0 < C < 1$

R = Rank below which C of all j^{th} value in sample size N should occur

$$p_i = \frac{N+1-i}{i} \frac{R}{1-R} p_{i-1} \qquad 1 \le i \le j-1,\ p_0 = 1$$

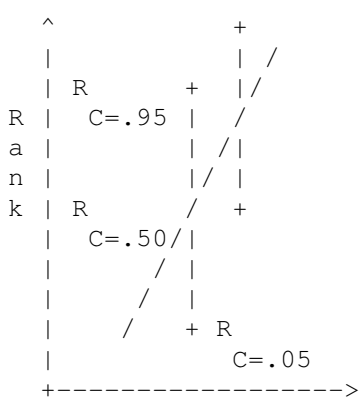
NOTE: $R_{C=.50} \cong (j - .3)/(N + .4)$ = median rank

$$R_{j|N,C=z} = 1 - R_{(N+1-j)|N,C=1-z}$$

e.g. $R_{3|10,C=.95} = .507 = 1 - R_{8|10,C=.05} = 1 - .493$

Following is an example with sample size N = 10 . A 'best-fit line' is calculated using maximum likelihood parameters. The confidence bounds are determined by plotting the Confidence Rank, CR, values versus the theoretical data value points as they would lie on the best-fit line. A confidence bound curve is drawn through these points.

For example, for N = 10



j	$R_{C=.05}$	$R_{C=.50}$	$R_{C=.95}$
1	.005	.067	.259
2	.036	.163	.394
3	.087	.260	.507
4	.150	.356	.607
5	.222	.452	.696
6	.305	.548	.778
7	.393	.644	.850
8	.493	.740	.913
9	.606	.837	.963
10	.741	.933	.995

References

1. Kapur, K.C. and L.R. Lamberson (1977), Reliability in Engineering Design, Wiley, New York.

2. Cohen, A. Clifford, Jr. (1963), 'Progressively Censored Samples in Life Testing', Technometrics, Vol. 5, No. 3, August, pp. 327-339.

3. Gajjar, A.V. and C.G. Khatri (1969), 'Progressively Censored Samplesd from Log-Normal and Logistic Distributions', Technometrics, Vol. 11, No. 4, November, pp. 793-803.

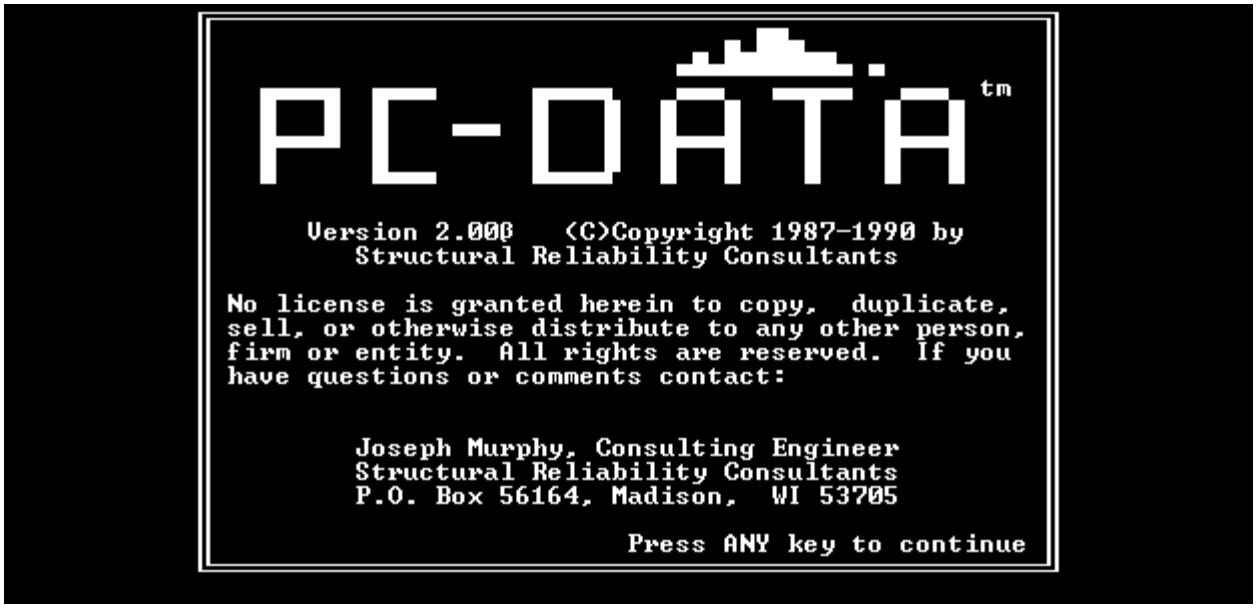
4. NASA Technical Support Package (1982), 'Consistent Tolerence Bounds for Statistical Distributions', MSC-20090, NASA Tech Briefs, Vol. 7, No. 1, Fall, 19 pages.

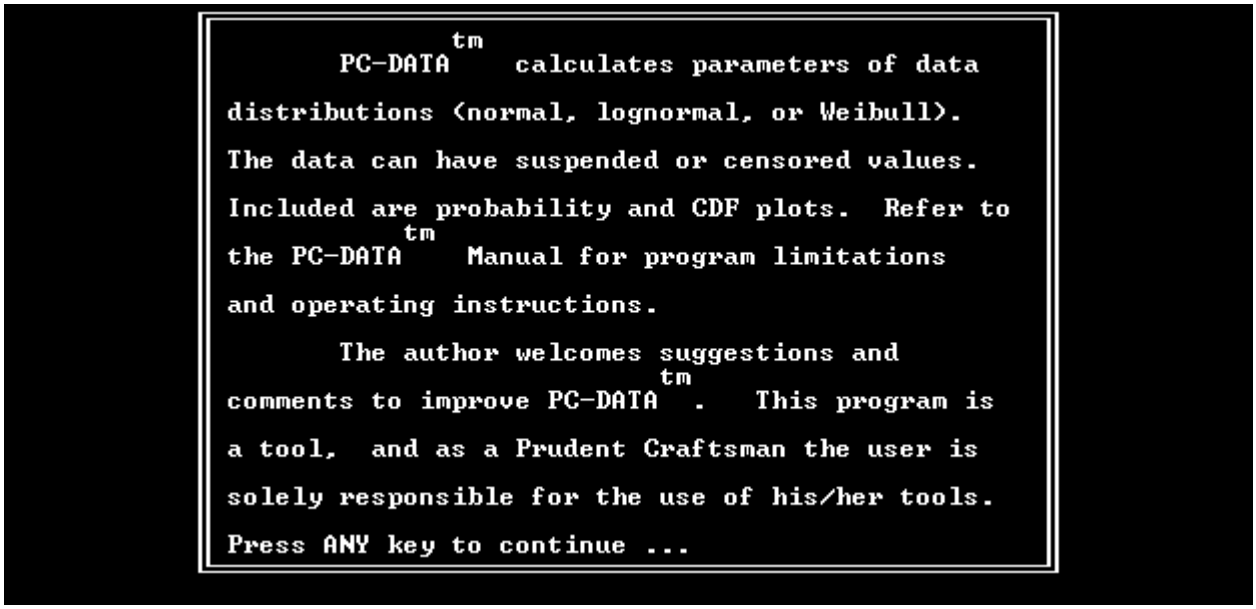
5. NASA Technical Support Package (1982), 'Computing Graphical Confidence Bounds', MSC-18908, NASA Tech Briefs, Vol. 7, No. 1, Fall, 10 pages.

6. Miller, Harry (1980), 'Statistical Fundamentals', Workshop Notes in Statistical Aspects of Testing of Composite Materials Workshop, sponsored by the Materials Laboratory of the Air Force Wright Aeronautical Laboratrories, Wright-Patterson Air Force Base, Ohio, 45433.

Input Screens

After two introduction screens follow two and maybe three input screens.





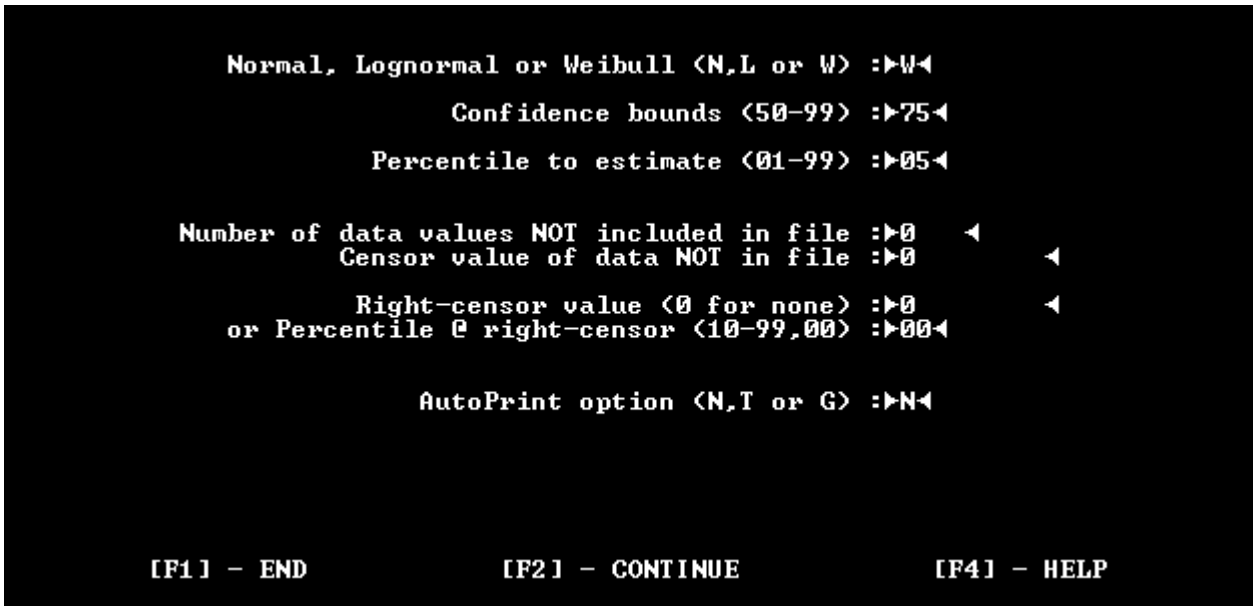
An example of the first input screen appears below:

If there are any files with extension *.DAT they will be listed. (If there are no files then this screen is skipped.)

The user chooses which one to analyze or to make a NEW file (99). If you know that the file has over 140 data points then add 100 to the file number (i.e., if SEX_FOX2 had over 140 points you would enter 104).

Enter 0 to return to DOS.

An example of the second input screen appears below:



The first field asks which distribution type for estimating parameters. Choose W,L or N.

The second field asks for the Binomial-Type Confidence bounds. Inputting 50 gives no bounds while inputting 75 gives the 75% and 25% confidence bounds about the maximum likelihood line.

The third field asks for the percentile to estimate at the lower confidence. For this example the program will calculate the fifth percentile with 75% confidence, x.05[75% CR].

The fourth field asks for the right censor value or suspension value. A nonzero entry here will essentially suspend all data above this value. This is useful for fitting 'lower tail' data. It will also be the maximum value plotted on the probability and CDF plots.

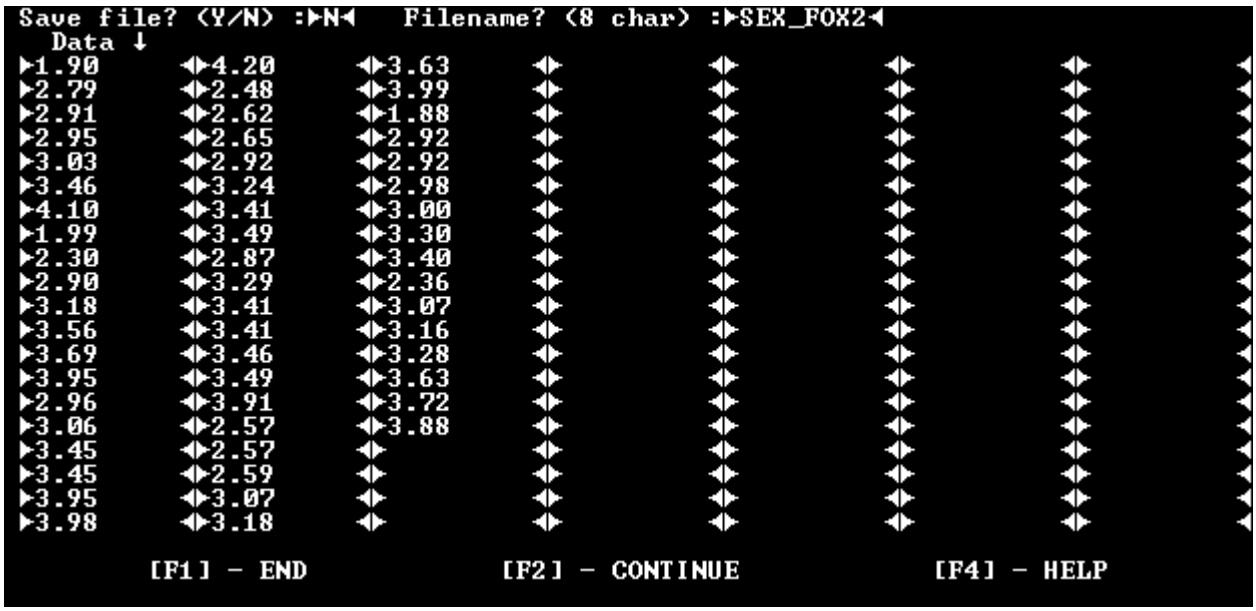
The fifth field asks for the number of data not included in the file. This is used with the fourth field to fit the 'tail'. If you proofload 100 items at level S and 10 fail, you can have a file with the 10 values, enter S as the right censor value and 90 as the number of data values NOT included in the file.

The last field on this screen is an autoprint option.
(Only works with Epson, IBM dot matrix printers.)
N for no automation,
T will print all text screens automatically, and
G will print all text and graphics screen automatically.

Pressing the

- F1 key will return you to DOS,
- F2 key will get you to the next screen, and
- F4 key will pop up a help window.

(The third input screen would be skipped if the file number entered in the first screen was greater than 100.) An example of the third input screen appears below:



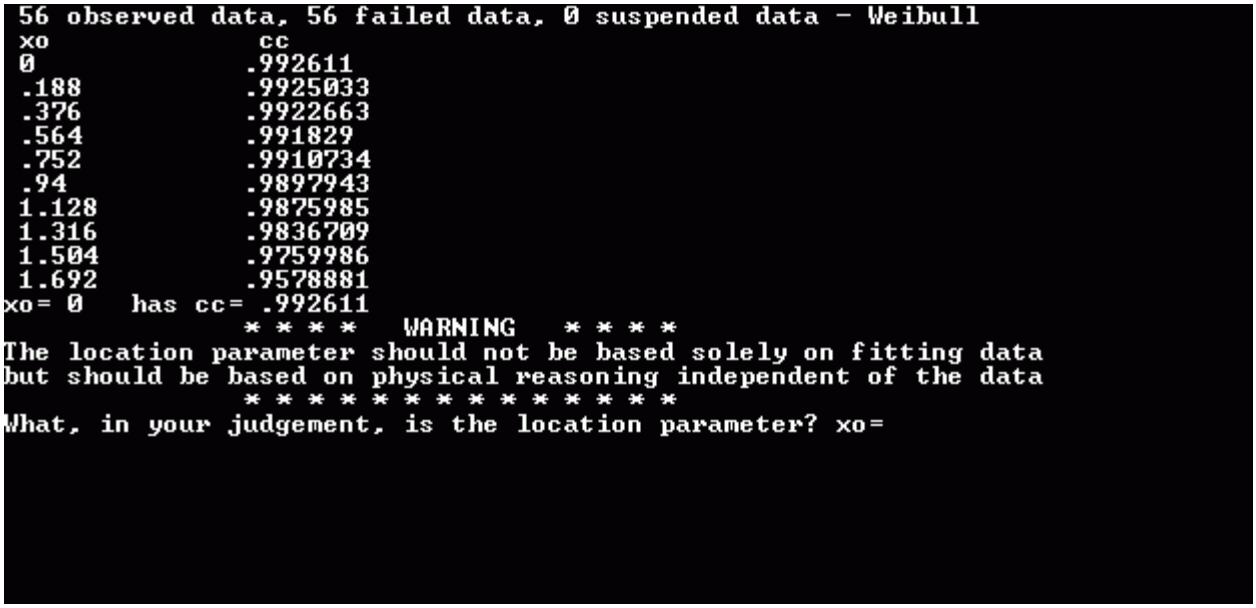
The first field asks if you want to save the file. If you do not change anything, you do not have to save the file. If you save the file and do not change the filename, the old file will be written over.

The second field asks for the filename by which to save the data.

The next 140 fields are data fields. They would be blank for a new file. One enters a negative value for suspended data. Again press the F4 key for help with editing.

Output Screens

There a three output screens. An example of the first output screen appears below:



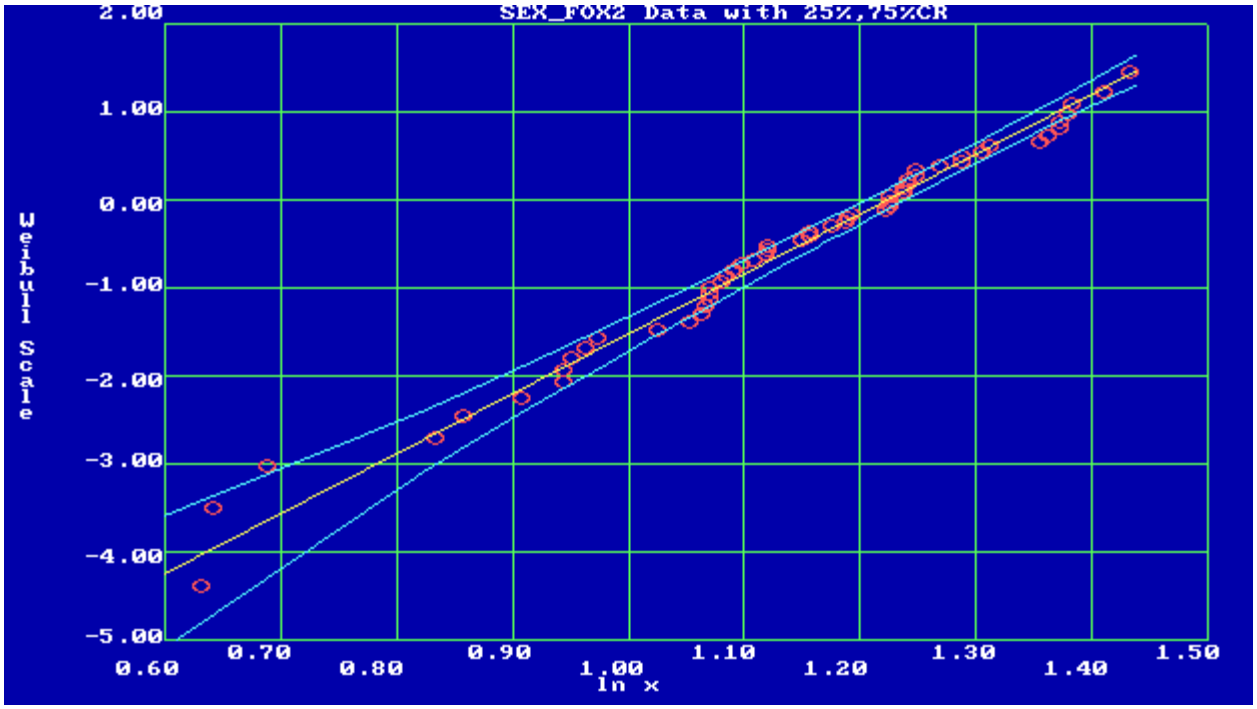
The first line gives the data number observed, failed and suspended and the distribution being fitted. If the distribution has a location parameter then ten estimates of the location parameter are calculated as 0, 0.1, 0.2, ..., 0.9 times the lowest ordered value. The correlation coefficient of the least squares best-fit straight line on distribution probability 'paper' are calculated. The final x_o used is the best 'quadratic' fit over the 10 location parameters and their correlation coefficients. REMEMBER the location parameter is not an estimate but an ABSOLUTE MINIMUM value.

After you choose the minimum value PC-DATA™ calculates distribution parameters by graphical and maximum likelihood techniques. The last line has the specified percentile calculated at specified confidence and 'best-fit' (i.e., 50%) confidence.

```
 56 observed data, 56 failed data, 0 suspended data - Weibull
 xo              cc
 0              .992611
 .188           .9925033
 .376           .9922663
 .564           .991829
 .752           .9910734
 .94            .9897943
 1.128          .9875985
 1.316          .9836709
 1.504          .9759986
 1.692          .9578881
xo= 0   has cc= .992611
              * * * *   WARNING   * * * *
The location parameter should not be based solely on fitting data
but should be based on physical reasoning independent of the data
              * * * * * * * * * * * * * * *
What, in your judgement, is the location parameter? xo= 0
xo= 0 by GRAPH    shape= 6.691522  scale(=x.63)= 3.393342  cc= .992611
        xmean= 3.166753  sigx= .5547987  covx= .1751948
xo= 0 by MAXLLHD  shape= 6.81671  scale(=x.63)= 3.393099
 x.47 = xmean= 3.169629  sigx= .5458433  covx= .1722105
 x.05[75% CR] = 2.082343     x.05[50% CR] = 2.194646
Press ENTER key to continue...
```

If you do not have graphics capabilities entering a 0 before pressing the enter key will return you to input screen 1.

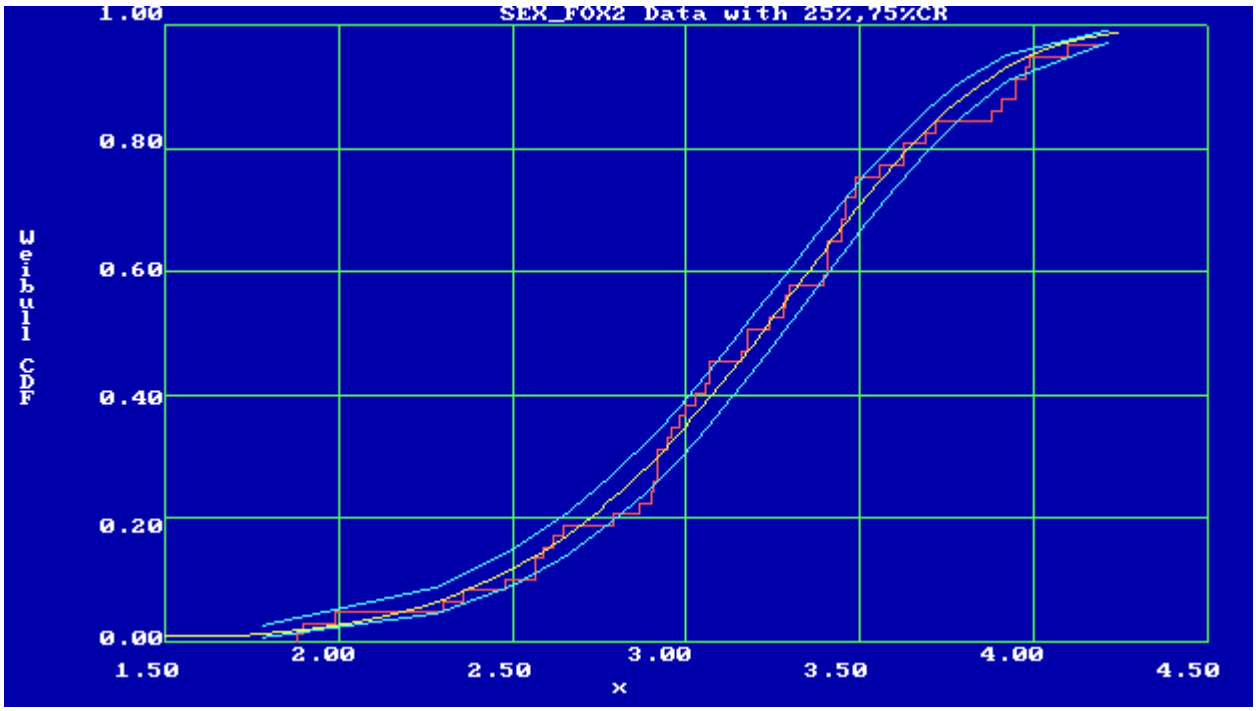
An example of the second output screen appears below:



This is a graphical plot of the data, fitted distribution and confidence bounds on appropriate probability axes. For this example one would expect 75% of the data to be to the right of the left confidence bound and 25% of the data to be to the right of the right confidence bound or 50% between the bounds.

To leave this screen press the [Esc] (escape) key. For help press the F4 key.

An example of the third output screen appears below:



This screen is a graphical plot of the CDF, both empirical and fitted. The horizontal axis is always the units of the data.

To leave this screen press the [Esc] (escape) key. For help press the F4 key. Leaving this screen returns you to input screen 1.