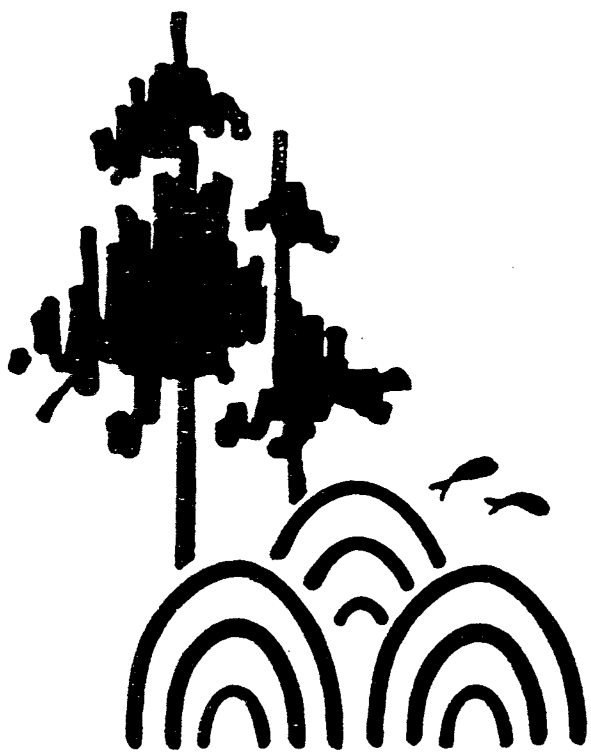
# WEIBUL

# A Program to Estimate Parameters of Forms of the Weibull Distribution Using Complete, Censored, and Truncated Data



Publication No. FWS-3-82
School of Forestry and Wildlife Resources
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
1982

# WEIBUL

# A Program to Estimate Parameters of Forms of the Weibull Distribution Using Complete, Censored, and Truncated Data

Bruce R. Zutter
Richard G. Oderwald
Robert M. Farrar, Jr.
Paul A. Murphy

Publication No. FWS-3-82

School of Forestry and Wildlife Resources
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

July, 1982

## ACKNOWLEDGMENTS

This study was funded in part by the Southern Forest Experiment Station, USDA Forest Service, under Cooperative Agreement No. 19-80-363.

## AUTHORS

The authors are, respectively, Research Associate, Auburn University, Auburn, AL; Associate Professor, VPI & SU, Blacksburg, VA; and Principal Mensurationists, Southern Forest Experiment Station, Forest Service - USDA, Monticello, AR. At the time of this study B. R. Zutter was a Graduate Research Assistant, Dept. of Forestry, VPI & SU.

TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF APPENDICES

# INTRODUCTION

The Weibull distribution has been used in many areas of forestry research. Complete Weibull distributions have been used to describe individual tree mortality (Glover and Hool 1979), stand level mortality (Somers *et al*. 1980), seed germination (Bonner and Dell 1976, Rink *et al*. 1979), and tree height and diameter distribution in forest stands (Bailey and Dell 1973, Hafley and Schreuder 1977, and others). Computer programs to estimate parameters of the complete Weibull distribution when the data are complete are readily available (Bailey 1974, Schreuder *et al*. 1978).

Sample data for some applications may be censored or truncated rather than complete. The program WEIBUL was prepared to allow estimation of complete or truncated Weibull distribution parameters from several types of censored or truncated data. Parameters can be estimated for the following Weibull distribution forms and data types:

1. complete distributions

    2 and 3 parameter using

    - complete data
    - left censored data
    - right censored data
    - left and right censored data (doubly censored)

2. truncated distributions

   2 parameter left truncated using

      left truncated data

      left truncated and right censored data

   2 and 3 parameter right truncated using

      right truncated data

      left censored and right truncated data

   2 parameter left and right truncated using

      left and right truncated data (doubly truncated)

Probability density functions and cumulative distribution functions for the above distribution types are given in Table 1.

WEIBUL is written in FORTRAN using double precision and contains approximately 1300 lines. It does not require any subroutines or functions other than those commonly available, such as square root or arcsine.

WEIBUL is available as a list, on cards, or on tape. Those requiring a tape copy should provide a tape and the necessary labeling information. Questions and requests regarding WEIBUL should be sent to

Department of Forestry
VPI & SU
Blacksburg, VA 24061.

Table 1. Probability density functions (pdf) and cumulative distribution functions (cdf) for forms of the Weibull distribution contained in program WEIBUL*

| Distribution form | Type | Equation |
|---|---|---|
| 3 parameter complete | pdf | $f(x) = (c/b)((x-a)/b)^{c-1}\exp(-((x-a)/b)^c)$ |
| | cdf | $F(x) = 1 - \exp(-((x-a)/b)^c)$ |
| | | $x \geq a \quad b, c > 0$ |
| 2 parameter complete | pdf, cdf | same as 3 parameter except a equals 0 and $x \geq 0$ |
| 2 parameter left truncated | pdf | $f(x) = (c/b)(x/b)^{c-1}\exp((t/b)^c - (x/b)^c)$ |
| | cdf | $F(x) = 1 - \exp((t/b)^c - (x/b)^c)$ |
| | | $x \geq t \quad b, c > 0$ |
| | | where |
| | | $t$ = left truncation point |
| 3 parameter right truncated | pdf | $f(x) = \dfrac{(c/b)((x-a)/b)^{c-1}\exp(-((x-a)/b)^c)}{1 - \exp(-(T-a)/b)^c}$ |
| | cdf | $F(x) = \dfrac{1 - \exp(-((x-a)/b)^c)}{1 - \exp(-((T-a)/b)^c)}$ |
| | | $a \leq x \leq T \quad b, c > 0$ |
| | | where |
| | | $T$ = right truncation point |
| 2 parameter right truncated | pdf, cdf | same as 3 parameter right truncated except a equals 0 and $0 \leq x \leq T$ |

Table 1. (continued)

| Distribution form | Type | |
|---|---|---|
| 2 parameter left and right truncated | pdf | $f(x) = \dfrac{(c/b)\ (x/b)^{c-1}\ \exp((t/b)^c - (x/b)^c)}{1 - \exp(-(T/b)^c)}$ |
| | cdf | $F(x) = \dfrac{1 - \exp((t/b)^c - (x/b)^c)}{1 - \exp(-(T/b)^c)}$ |
| | | $t \leq x \leq T \quad b,\ c > 0$ |

* a, b, and c are, respectively, the location, scale, and shape parameters.

# PROGRAM DESCRIPTION

## Input

Data may be either ungrouped or grouped. A maximum of 500 observations for ungrouped data or 500 class midpoints for grouped data may be entered for any single data set. There is no limit on the number of data sets that may be entered for a single program execution. On input a check is made for errors dealing with mis-specification of data type. If such an error occurs, an error code is printed and processing of that data set is terminated. A list of input error codes and their description is given in Appendix II.

Data are entered through the user supplied subroutine DATIN. The subroutine name and argument list are:

```
SUBROUTINE DATIN (XD,NXD,IOUT,IGRPT,IDIS,GRP1,TLT,TRT,
  CLT,CRT,NCL,NPUN,NFIT,TITLE)
```

All arguments must appear and in the order given above even if the argument values are blank. Descriptions of the arguments are given in Table 2. All real variables in DATIN should be double precision. An example of DATIN is given in Appendix Ia.

An end-of-data check must appear in DATIN, and the check value must be put at the end of each data set. An example data set to be read by DATIN is given in Appendix Ib.

Table 2. Description of variables used in subroutine DATIN

| Variable | Type | Comments |
|---|---|---|
| XD | real | array of length $\leq$ 500 of class midpoints if the data are grouped or individual observations if the data are not grouped |
| NXD | integer | array of length $\leq$ 500 of class frequencies for grouped data (leave blank for ungrouped data) |
| NUM | integer | number of class midpoints or individual observations in the data set |
| IOUT | integer | output code giving the type of output desired<br>1 = printed output only<br>2 = printed and punched output<br>3 = punched output only<br>(optional, default = 1) |
| IGRPT | integer | identifies input data grouping<br>0 = ungrouped data<br>1 = grouped data |
| IDIS | integer | identifies distribution and data type<br>0 = 3 parameter complete distribution<br>complete data<br>1 = 2 parameter complete distribution<br>complete data<br>2 parameter left truncated distribution<br>left truncated data<br>2 = 3 parameter complete distribution<br>left censored data<br>right censored data<br>left and right censored data<br>3 parameter right truncated distribution<br>right truncated data<br>left censored and right truncated data |

Table 2. (continued)

| Variable | Type | Comments |
|---|---|---|
| IDIS (continued) | integer | 3 = 2 parameter left truncated distribution<br>left truncated and right censored data<br>2 parameter right truncated distribution<br>right truncated data<br>left censored and right truncated data<br>2 parameter complete distribution<br>left censored data<br>right censored data<br>left and right censored data<br>2 parameter left and right truncated distribution<br>left and right truncated data<br><br>(optional, default = 0) |
| GRP1 | real | desired class interval width for distribution summary and statistic calculation<br><br>(optional, default = 1.) |
| TLT | real | left truncation point (optional) |
| TRT | real | right truncation point (optional) |
| CLT | real | left censoring point (optional) |
| CRT | real | right censoring point (optional) |
| NCL | integer | number of left censored observations (optional) |
| NCR | integer | number of right censored observations (optional) |
| NPUN | integer | number of unit to which punched ouput is to be sent if IOUT $\geq$ 2<br><br>(optional, default = 7) |
| NFIT | integer | total number of data sets to be fit |
| TITLE | real | array of length 10 which contains an 80-character title in elements of length 8<br><br>(optional, default is blanks) |

Output

Program results may be printed and/or punched. An example of printed results is given in Appendix Ic. The format of punched results is given in Table 3. Punched results may be directed to cards, tape, or disk upon appropriate specification of the value of input variable NPUN (see Table 2).

Printed results for each data set consists of: the title (if any), data set number, distribution and data type; mean, standard deviation, variance, minimum, maximum, and total number of the observations; truncation and/or censoring values and the number of censored observations (if any); estimated parameter values and fit statistics; a summary by class interval of observed, predicted, and residual proportions, observed and predicted cumulative proportions, and observed, predicted, and relative frequency. For censored data only predicted proportions and frequencies are given for each class interval in the censored region, while observed, predicted and residual frequencies are given for the entire censored region.

The fit statistics are the chi-square and Kolmogorov-Smirnov statistics, and the sum of the absolute deviations (SABS) of the observed from the predicted distribution. All three fit statistics are computed using observed and predicted frequencies for the printed class intervals with the entire censored region (if any) comprising one interval.

Table 3. Format of punched output from WEIBUL

| Columns | Format | Description |
|---|---|---|
| 1-3 | I3 | data set number |
| 4-5 | I2 | condition code for fit<br>0 = specified distribution fit<br>1 = 2 parameter Weibull fit instead of specified 3 parameter Weibull<br>2 = data set could not be fit |
| 6-15 | F10.6 | "a" parameter |
| 16-25 | F10.6 | "b" parameter |
| 26-35 | F10.6 | "c" parameter |
| 36-43 | F8.4 | left truncation point (0.0 if no left truncation) |
| 44-51 | F8.4 | right truncation point (0.0 if no right truncation) |
| 52-59 | F8.4 | left censoring point (0.0 if no left censoring) |
| 60-67 | F8.4 | right censoring point (0.0 if no right censoring) |
| 68-72 | F5.0 | number of left censored observations |
| 73-77 | F5.0 | number of right censored observations |

Punched results consist of the data set number, a code indicating if the desired distribution has been fit (see Table 3), estimated Weibull parameters, truncation points (if any), censoring points (if any), and the number of left and right censored observations (if any).

## PARAMETER ESTIMATION METHOD

Joint maximum likelihood estimation is used in solving for the values of the Weibull parameters. Initial guesses or starting values for the parameters are found and the partial derivatives (maximum likelihood equations) of the log-likelihood equation simultaneously solved using an iterative process.

### Initial Values for Parameters

The Berger and Lawrence (1974) method of estimating Weibull parameters using linear regression is used in WEIBUL to supply initial starting values for the shape and scale parameters given the initial value of the location parameter. For a 3 parameter distribution fit with complete (left censored) data two initial values of the location parameter are calculated as the trisection points of the interval from zero to the first order statistic (zero to the left censoring point). For left truncated distributions fit with left truncated data and for 2 parameter distributions fit with complete or censored data the location parameter is zero.

For all data and distribution types the data points are put in ascending order and a cdf value for each point is computed based on the ordered data. The 3 parameter complete distribution is linearized as

$$\ln(\ln(1/(1-F(x)))) = c\ln(x-a) - c\ln b$$

where $F(x)$ denotes the cdf value at x, and ln denotes the natural logarithm. The values of x, $F(x)$, and the initial value of the location parameter (a) are inserted into the linearized cdf function, and c and clnb are estimated as regression coefficients. The initial value of the scale parameter (b) is then recovered.

When two initial values of the location parameter (a) are found, a set of b and c are calculated for each of the a's. Only one set of initial values are used if the subsequent iteration procedure converges. If the procedure fails to converge with the first set of initial values, the second set is used.

Maximum Likelihood Equation Solutions

Solutions of the maximum likelihood estimates of the parameters are obtained using an iterative process given by Wingo (1972, 1973). The process, known as constrained modified quasilinearization (Wingo 1973) searches for parameter values over a bounded solution space. Lower bounds in WEIBUL for the a, b and c parameters are $1 \times 10^{-15}$, .1, and .1, respectively, for the 3 parameter Weibull distribution. The upper bound on the location parameter is set equal to slightly less than the first order statistic of the sample (or the left censoring point if left censoring occurs). The upper bounds on the b and c parameters are 100.0 and 20.0, respectively.

The iterative search assures an improvement in the solution with each iteration in terms of the value of a performance index, P. The performance index equals the product of the vector of values of the maximum likelihood equations and its transpose. After each iteration P is computed and when P is less than $1 \times 10^{-20}$ the search is terminated. If a 3 parameter Weibull distribution has been specified and the first set of starting values for the parameters does not lead to a solution, then the iterative search for the parameters begins again using the second set of starting values. If the second set of values does not lead to a solution a 2 parameter distribution is specified and the search for the parameter values continues. If a 2 parameter distribution cannot be fit a message indicating that no distribution could be fit to the data is printed.

# LITERATURE CITED

Bailey, R. L. 1974. Computer programs for quantifying diameter distributions with the Weibull function. For. Sci. 20:229

Bailey, R. L. and T. R. Dell. 1973. Quantifying diameter distributions with the Weibull function. For. Sci. 19:97-104.

Berger, R. W. and K. Lawrence. 1974. Estimating Weibull parameters by linear and non-linear regression. Technometrics 16:617-619.

Bonner, F. T. and T. R. Dell. 1976. The Weibull function: a new method of comparing seed vigor. Jour. Seed Tech. 1:97-103.

Glover, G. R. and J. N. Hool. 1979. A basal area ratio predictor of loblolly pine plantation mortality. For. Sci. 25:275-282.

Hafley, W. L. and H. T. Schreuder. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Can. J. For. Res. 7:481-487.

Rink, G., T. R. Dell, G. Switzer, and F. T. Bonner. 1979. Use of the Weibull function to quantify sweetgum germination data. Silvae Genetica 28:9-12.

Somers, G. L., R. G. Oderwald, W. R. Harms, and O. G. Langdon. 1980. Predicting mortality with a Weibull distribution. For. Sci. 26:291-300.

Schreuder, H. T., W. L. Hafley, E. W. Whitehorne, and B. J. Dare. 1978. Maximum likelihood estimation of selected distributions (MLESD). School of For. Res. N. C. State U. Tech. Rep. No. 61. 21 p.

Wingo, D. R. 1972. Maximum likelihood estimation of the parameters of the Weibull distribution by modified quasilinearization. IEEE Trans. Rel. Vol. R-21(2):89-93.

Wingo, D. R. 1973. Solution of the three-parameter Weibull equations by constrained modified quasilinearization (progressively censored samples). IEEE Trans. Rel. Vol. R-22(2): 96-102.

Appendix Ia. Example Execution of WEIBUL: Subroutine DATIN.

This example is for reading grouped data left truncated at 3.5. The class midpoints are located in columns 1 through 7 and the class frequencies in columns 8 through 11 of the input cards (see Appendix Ib for the data). Only one data set is to be fit with both printed and punched output desired. A left truncated distribution is to be fit with the distribution summary to be given over intervals of 1 unit in width. A count of the number of data points is made as the data are read. The read terminates when a "dummy" midpoint having a negative value is read.

```
      SUBROUTINE DATIN(XD,NXD,NUM,IOUT,IGRPT,IDIS,GRP1,TLT,TRT,CLT,CRT,
     %NCL,NCR,NPUN,NFIT,TITLE)
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION XD(500),NXD(500),TITLE(10)
      NFIT=1
      IOUT=2
      IGRPT=1
      IDIS=1
      GRP1=1.D0
      TLT=3.5D0
      NPUN=7
      READ(5,15) (TITLE(I),I=1,10)
      DO 5 I=1,100
      READ(5,20) XD(I),NXD(I)
      IF(XD(I).LT.0.D0) GO TO 7
    5 NUM=NUM+1
    7 RETURN
   15 FORMAT(10A8)
   20 FORMAT(F6.0,1X,I4)
      END
```

Appendix Ib. Example Execution of WEIBUL: Input Data.

```
SAMPLE RUN OF PROGRAM WEIBUL
  4.     209.
  5.     132.
  6.      92.
  7.      66.
  8.      51.
  9.      53.
 10.      60.
 11.      63.
 12.      54.
 13.      54.
 14.      42.
 15.      16.
 16.       8.
 17.       1.
-100.
```

Appendix Ic. Example Execution of WEIBUL: Printed Results.

SAMPLE RUN OF PROGRAM WEIBUL

DATA SET NUMBER 1

TRUNCATED WEIBULL DISTRIBUTION - LEFT TRUNCATED DATA

INPUT DATA SUMMARY - GROUPED DATA

| | | | |
|---|---|---|---|
| MEAN OBS = | 7.7814 | MAX OBS = | 17.0000 |
| STD. DEV = | 3.4779 | MIN OBS = | 4.0000 |
| VARIANCE = | 12.0955 | | |

LEFT TRUNCATION POINT = 3.5000

TOTAL NO. OBS = 901.

FITTING SUMMARY

| | | |
|---|---|---|
| WEIBULL | | |
| PARAMETER | B = | 6.522468 |
| VALUES | C = | 1.492674 |

| | | |
|---|---|---|
| FITTING | SUM OF ABSOLUTE DEVIATIONS = | 306.3511 |
| STATISTICS | KOLMOGOROV-SMIRNOV = | 0.2320 |
| | CHI-SQUARE = | 163.0438 |

DISTRIBUTION SUMMARY

| LOWER AND UPPER BOUNDS ON INTERVAL | | OBS PROP | PRED PROP | RESID PROP | CUMUL OBS | CUMUL PRED | OBSERVED FREQUENCY | PREDICTED FREQUENCY | RESIDUAL FREQUENCY |
|---|---|---|---|---|---|---|---|---|---|
| 3.5000- | 4.5000 | 0.2320 | 0.1645 | 0.0675 | 0.2320 | 0.1645 | 209.000 | 148.227 | 60.773 |
| 4.5000- | 5.5000 | 0.1465 | 0.1519 | -.0054 | 0.3785 | 0.3164 | 132.000 | 136.871 | -4.871 |
| 5.5000- | 6.5000 | 0.1021 | 0.1348 | -.0326 | 0.4806 | 0.4512 | 92.000 | 121.414 | -29.414 |
| 6.5000- | 7.5000 | 0.0733 | 0.1158 | -.0425 | 0.5538 | 0.5669 | 66.000 | 104.305 | -38.305 |
| 7.5000- | 8.5000 | 0.0566 | 0.0968 | -.0402 | 0.6104 | 0.6638 | 51.000 | 87.227 | -36.227 |
| 8.5000- | 9.5000 | 0.0588 | 0.0791 | -.0203 | 0.6693 | 0.7428 | 53.000 | 71.259 | -18.259 |
| 9.5000- | 10.5000 | 0.0666 | 0.0633 | 0.0033 | 0.7358 | 0.8061 | 60.000 | 57.013 | 2.987 |
| 10.5000- | 11.5000 | 0.0699 | 0.0497 | 0.0202 | 0.8058 | 0.8558 | 63.000 | 44.761 | 18.239 |
| 11.5000- | 12.5000 | 0.0599 | 0.0383 | 0.0216 | 0.8657 | 0.8941 | 54.000 | 34.537 | 19.463 |
| 12.5000- | 13.5000 | 0.0599 | 0.0291 | 0.0308 | 0.9256 | 0.9232 | 54.000 | 26.221 | 27.779 |
| 13.5000- | 14.5000 | 0.0466 | 0.0218 | 0.0249 | 0.9723 | 0.9450 | 42.000 | 19.609 | 22.391 |
| 14.5000- | 15.5000 | 0.0178 | 0.0160 | 0.0017 | 0.9900 | 0.9610 | 16.000 | 14.457 | 1.543 |
| 15.5000- | 16.5000 | 0.0089 | 0.0117 | -.0028 | 0.9989 | 0.9727 | 8.000 | 10.515 | -2.515 |
| 16.5000- | 17.5000 | 0.0011 | 0.0084 | -.0073 | 1.0000 | 0.9811 | 1.000 | 7.550 | -6.550 |
| 17.5000- | 18.5000 | 0.0 | 0.0059 | -.0059 | 1.0000 | 0.9870 | 0.0 | 5.354 | -5.354 |
| 18.5000- | 19.5000 | 0.0 | 0.0042 | -.0042 | 1.0000 | 0.9912 | 0.0 | 3.753 | -3.753 |
| 19.5000- | | 0.0 | 0.0088 | -.0088 | 1.0000 | 1.0000 | 0.0 | 7.928 | -7.928 |

Appendix II. Input Error Codes for Misspecification of Data Types.

| Error Number | Description |
|---|---|
| 1 | left truncation and left censoring have been specified |
| 2 | right truncation and right censoring have been specified |
| 3 | the left truncation or censoring point is larger than a data point or class midpoint |
| 4 | the right censoring point is smaller than a data point or class midpoint |
| 5 | the right truncation point is smaller than a data point or class midpoint |