**Statistical Bulletin**
Reliability & Variation Research

DETROIT RESEARCH INSTITUTE
P.O. BOX 36504 • GROSSE POINTE, MICHIGAN 48236 • (313) 886-7976

LEONARD G. JOHNSON
EDITOR

WANG H. YEE
DIRECTOR

Volume 13 February , 1984

Bulletin 8 

---

# THE LOG-NORMAL COMPUTER PROGRAM (LGNPRG)

## INTRODUCTION

The Log-Normal Distribution has mathematical properties which make it useful in studying certain classes of phenomena related to ecology and the enviroment. For example, the log-normal distribution has been successfully applied to the following studies:

(a) Particle size variability
(b) Pollen count studies related to allergies
(c) Concentrations of toxic substances in water
(d) Automobile exhaust emission levels

The mathematical properties which make the log-normal distribution applicable in these studies are

(1) The fact that real logarithms exist only for positive numbers. As such, the log-normal does not have a tail to the left of X = 0.

(2) In order to increase the normal Z-Score at any X by an amount h it is necessary to multiply X by the factor $e^{h\sigma}$, where $\sigma$ is the standard deviation of lnX. This implies that the largest values in the distribution can be enormous multiples of average values, as is true of unfiltered particles and gross cases of pollution. Incidentally, this property of the log-normal distribution makes it unfit for fatigue life studies, since it implies a decreasing hazard rate with respect to exposure time at the high end of the distribution.

## DEFINITION OF A LOG-NORMALLY DISTRIBUTED VARIABLE

A variable X is said to be log-normally distributed if its natural logarithm (lnX) has a normal (Gaussian) distribution. From this definition it follows that the probability density function f(X) of a log-normally distributed variable X is

$$f(X) = \frac{1}{X \cdot \sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln X - M}{\sigma}\right)^2} \qquad (1)$$

where,

X = the log-normally distributed variable

M = the mean of the natural logarithm of X

$\sigma$ = the standard deviation of the natural logarithm of X

## THE ESTIMATION OF M AND $\sigma$

The estimation of M and $\sigma$ of (1) from a sample of N observations on the variable X is straightforward. We simply take the natural logarithms of the N observations $(X_1, X_2, \ldots, X_N)$, and thus generate the set $(\ln X_1, \ln X_2, \ldots, \ln X_N)$.

Then ,

$$M = \frac{\ln X_1 + \ln X_2 + \ldots + \ln X_N}{N} \qquad (2)$$

and

$$\sigma = \sqrt{\frac{(M - \ln X_1)^2 + (M - \ln X_2)^2 + \ldots + (M - \ln X_N)^2}{N - 1}} \qquad (3)$$

## LOCATING THE NINE DECILES OF THE DISTRIBUTION

**DEFINITION:** The nine deciles of a distribution of a variable X are defined to be the values of the variable X below which are found, respectively, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, and 90 % of the population.

To locate these deciles in a log-normal population we simply use NORMAL Z-SCORES on the natural logarithms of the variable X.

Thus,

10 % location of $\ln X = M - 1.282155\sigma$

20 % location of $\ln X = M - 0.84162\sigma$

30 % location of $\ln X = M - 0.52440\sigma$

40 % location of $\ln X = M - 0.25335\sigma$

50 % location of $\ln X = M$

60 % location of $\ln X = M + 0.25335\sigma$

70 % location of $\ln X = M + 0.52440\sigma$

80 % location of $\ln X = M + 0.84162\sigma$

90 % location of $\ln X = M + 1.282155\sigma$

In the above expressions, $M$ = Mean of $\ln X$

$\sigma$ = Standard Deviation of $\ln X$

By taking anti-logs (exponentials) of these values of ln X we obtain the nine deciles of X as the following:

10 % location of X = $EXP(M - 1.282155\sigma)$

20 % location of X = $EXP(M - 0.84162\sigma)$

30 % location of X = $EXP(M - 0.52440\sigma)$

40 % location of X = $EXP(M - 0.25335\sigma)$

50 % location of X = $EXP(M)$

60 % location of X = $EXP(M + 0.25335\sigma)$

70 % location of X = $EXP(M + 0.52440\sigma)$

80 % location of X = $EXP(M + 0.84162\sigma)$

90 % location of X = $EXP(M + 1.28155\sigma)$

In a similar fashion, we locate the 1 %, 5 %, 95 %, and 99 % levels of X as follows:

1 % location of X = $EXP(M - 2.32635\sigma)$

5 % location of X = $EXP(M - 1.64485\sigma)$

95 % location of X = $EXP(M + 1.64485\sigma)$

99 % location of X = $EXP(M + 2.32635\sigma)$

## SAMPLING ERRORS OF ESTIMATED QUANTILE LEVELS

**DEFINITION:** The quantile level Q of a variable X is that value of X below which is found the fraction Q of the population.

Thus for the nine deciles we have

Q = .1, .2, .3, .4, .5, .6, .7, .8, and .9, respectively.

| | |
|---|---|
| For the $1^{st}$ percentile : | Q = .01 |
| For the $5^{th}$ percentile : | Q = .05 |
| For the $95^{th}$ percentile: | Q = .95 |
| For the $99^{th}$ percentile: | Q = .99 |
| Etc. | Etc. |

For a NORMALLY distributed variable the standard error of quantile Q of the variable, as estimated from a sample of size N of the variable, is

$$\frac{\sigma}{Y_Q}\sqrt{\frac{Q(1 - Q)}{N}} \quad *$$

where, $\sigma$ = standard deviation of the variable

$Y_Q$ = Std. NORMAL ORDINATE at quantile Q (From a table of Gaussian ordinates.)

* See DUNLAP & KURTZ: Handbook of Statistical Charts and Tables World Book Co. (1932) Formula No. 431, pg. 140.

Hence, assuming the SAMPLE QUANTILE Q is also normally distributed, we can take the two-sided 90 % CONFIDENCE BAND of the ESTIMATED QUANTILE Q of the variable to be

$$\begin{pmatrix}\text{ESTIMATED QUANTILE Q}\\ \text{OF NORMAL VARIABLE}\end{pmatrix} \pm \frac{1.64485\,\sigma}{Y_Q}\sqrt{\frac{Q(1-Q)}{N}}$$

For a log-normal variable X, we have

$\sigma$ = Population Standard Deviation of ln X. [Use estimate (3) on page 3 if no better value is available]

For different Q's the estimated quantiles of ln X are given on page 4.

The Gaussian Ordinates $Y_Q$ are as follows:

| Q | $Y_Q$ |
|---|---|
| .01 | .02665 |
| .05 | .10314 |
| .10 | .17550 |
| .20 | .28000 |
| .30 | .34769 |
| .40 | .38634 |
| .50 | .39894 |
| .60 | .38634 |
| .70 | .34769 |
| .80 | .28000 |
| .90 | .17550 |
| .95 | .10314 |
| .99 | .02665 |

Hence, at quantile Q, the LOWER 5 % BOUNDARY of a log-normal variable X is

$$X_L = \text{EXP}\left[ M + Z_Q\sigma - \frac{1.64485\,\sigma}{Y_Q}\sqrt{\frac{Q(1-Q)}{N}} \right] \qquad (4)$$

Likewise, the UPPER 95 % BOUNDARY of the same log-normal variable X is (at quantile Q)

$$X_U = \text{EXP}\left[ M + Z_Q\sigma + \frac{1.64485\,\sigma}{Y_Q}\sqrt{\frac{Q(1-Q)}{N}} \right] \qquad (5)$$

In the above formulas: $Z_Q$ = NORMAL Z-SCORE for quantile Q

## THE LOG-NORMAL PROGRAM

In the log-normal computer program (LGNPRG) listing which follows, we use (2) and (3) to estimate the mean and sigma of the natural logs. The CENTRAL VALUES of PERCENTILES are obtained from the formulas on page 5. Finally, LOWER 5 % and UPPER 95 % bounds on percentiles are obtained by using (4) and (5) above.

The data are entered in statement 2 of the program. If the data are not grouped, simply enter

2 DATA N, $X_1, X_2, \ldots, X_N$ (Here N = Sample Size; $X_i$ = ith observation of the sample)

If the data are grouped, enter

2 DATA G,N,$X_1, X_2, \ldots, X_N, f_N$ (N = Sample Size; G = Number of Groups; $X_i$ = Central X in Group i; $f_i$ = Frequency in Group i)

It should be noted that the program asks whether or not a log-normal fit is desired, and whether or not the data are grouped. (Answer by a ZERO (0) for "NO", or by a 1 for "YES" in each case, and press the carriage return.). In case a "NO" answer is given to the log-normal fit question, the program simply gives an ordinary NORMAL CURVE FIT to the data.

## LOG-NORMAL PROGRAM (LGNPRG)

```
0 DIM X(300),H(300)
1 DIM T(10),Y(10),S(10),U(10),L(10),V(10),A(10),W(10),B(10)
2 DATA 6,100,1.5,10,3.5,20,6,30,8,25,10.5,10,13.5,5
8 PRINT"DØ YØU WANT A LØG-NØRMAL FIT(0=NØ, 1=YES)";
9 INPUT Z9
10 PRINT"IS DATA GRØUPED(0=NØ, 1=YES)";
11 INPUT Z8
12 IF Z8>.01 THEN 16
13 READ N
14 LET G = N
15 GØ TØ 17
16 READ G,N
17 FØR I = 1 TØ G
18 IF Z8>.01 THEN 23
21 READ X(I)
22 GØ TØ 24
23 READ X(I),H(I)
24 IF Z9>.01 THEN 27
25 NEXT I
26 GØ TØ 29
27 LET X(I)=.4342945*LØG(X(I))
28 GØ TØ 25
29 FØR I = 1 TØ G
30 IF Z8>.01 THEN 33
31 LET S1 = S1+X(I)
32 GØ TØ 35
33 LET S1 = S1 +H(I)*X(I)
34 GØ TØ 37
35 LET S2 = S2+X(I)*X(I)
36 GØ TØ 38
37 LET S2 = S2+H(I)*X(I)*X(I)
38 NEXT I
39 LET M = S1/N
40 LET D = SQR((N*S2-S1*S1)/(N*(N-1)))
42 LET T(1)=.25335
44 LET T(2)=.5244
46 LET T(3)= .84162
48 LET T(4)= 1.28155
50 LET T(5)= 1.64485
52 LET T(6)= 2.32635
55 LET Y(1)= 1.268
60 LET Y(2)= 1.318
70 LET Y(3)= 1.4288
80 LET Y(4)= 1.7094
90 LET Y(5)= 2.1137
100 LET Y(6)= 3.7331
110 FØR I = 1 TØ 6
120 LET S(I) = (Y(I)*D)/SQR(N)
130 NEXT I
140 FØR I = 1 TØ 6
150 LET L(I) = M-D*T(I)
160 LET U(I) = M+D*T(I)
```

```
170 NEXT I
180 FØR I = 1 TØ 6
190 LET V(I) = L(I) -S(I)*T(5)
200 LET W(I) = L(I) + S(I)*T(5)
210 NEXT I
220 FØR I = 1 TO 6
230 LET A(I) = U(I) -S(I)*T(5)
240 LET B(I) = U(I) + S(I)*T(5)
250 NEXT I
260
280 IF Z9>.01 THEN 630
290 PRINT"CENTRAL VALUES ØF LØWER PERCENTILES"
300 PRINT
310 PRINT"1ST","10TH","20TH","30TH","40TH"
320 PRINT L(6),L(4),L(3),L(2),L(1)
330 PRINT
340 PRINT
350 PRINT"CENTRAL VALUES ØF UPPER PERCENTILES"
360 PRINT
370 PRINT "60TH","70TH","80TH","90TH","99TH"
380 PRINT U(1),U(2),U(3),U(4),U(6)
390 PRINT
400 PRINT
410 PRINT"LØWER 5 % BØUNDS ØN LØWER PERCENTILES"
420 PRINT
430 PRINT"1ST","10TH","20TH","30TH","40TH"
440 PRINTV(6),V(4),V(3),V(2),V(1)
450 PRINT
459 PRINT
470 PRINT"LØWER 5 % BØUNDS ØN UPPER PERCENTILES"
480 PRINT
490 PRINT"60TH","70TH","80TH","90TH","99TH"
500 PRINT A(1),A(2),A(3),A(4),A(6)
510 PRINT
520 PRINT
530 PRINT"UPPER 95 % BØUNDS ØN LØWER PERCENTILES"
535 PRINT
540 PRINT"1ST","10TH","20TH","30TH","40TH"
550 PRINT W(6),W(4),W(3),W(2),W(1)
560 PRINT
570 PRINT
580 PRINT"UPPER 95 % BØUNDS ØN UPPER PERCENTILES"
590 PRINT
595 PRINT"60TH","70TH","80TH","90TH","99TH"
600 PRINT B(1),B(2),B(3),B(4),B(6)
605 IF Z9>.01 THEN 621
610 PRINT
612 PRINT"MEAN=";M
614 PRINT"SIGMA=";D
616 PRINT
```

```
620  GO TO 720
621 PRINT"MEAN OF COMMON LØGS=";M
622 PRINT"SIGMA OF COMMØN LØGS=";D
623 LET R9=2.3025851*M
624 LET R8 = 2.3025851*D
625 PRINT"MEAN OF NATURAL LØGS=";R9
626 PRINT"SIGMA OF NATURAL LØGS=";R8
627 LET M=10↑M
628 PRINT"50TH PERCENTILE=";M
629 GO TO 750
630 FOR I = 1 TO 6
640 LET L(I) = 10↑L(I)
650 LET U(I) = 10↑U(I)
660 LET V(I) = 10↑V(I)
670 LET A(I) = 10↑A(I)
680 LET W(I) = 10↑W(I)
690 LET B(I) = 10↑B(I)
700 NEXT I
710 GØ TO 290
720 PRINT
725 PRINT
730 LET R7 = M +(1.645*D)/SQR(N)
735 PRINT"UPPER 95 % LIMIT ØN MEAN =";R7
740 STOP
750 PRINT
755 PRINT
760 LET R6 = EXP(R9 +(1.645*R8)/SQR(N))
765 PRINT"UPPER 95% LIMIT ON INVERSE TRANSFØRM ØF MEAN ØF LØGS=";R6
790 LET A9 = EXP(R9+.5*R8*R8)
800 LET A8 = A9*SQR(-1 + EXP(R8*R8))
810 PRINT
820 PRINT
830PRINT"MEAN ØF ØRIG. VARIABLE(ASSUMING IT IS LØG-NØRMAL)=";A9
840 PRINT
850PRINT"SIGMA ØF ØRIG. VARIABLE(ASSUMING IT IS LØG-NØRMAL)=";A8
900 END

READY
```

## LOG-NORMAL EXAMPLE

```
2 DATA 6,100,1.5,10,3.5,20,6,30,8,25,10.5,10,13.5,5
```

(100 data points divided up into 6 groups)

RUN

DØ YØU WANT A LØG-NØRMAL FIT(0=NØ,1=YES) ? 1
IS DATA GRØUPED(0=NØ,1=YES) ? 1
CENTRAL VALUES OF LØWER PERCENTILES

| 1ST | 10TH | 20TH | 30TH | 40TH |
|---|---|---|---|---|
| 1.45135 | 2.65085 | 3.41618 | 4.10176 | 4.79556 |

CENTRAL VALUES ØF UPPER PERCENTILES

| 60TH | 70TH | 80TH | 90TH | 99TH |
|---|---|---|---|---|
| 6.42264 | 7.50901 | 9.01597 | 11.619 | 21.2216 |

LØWER 5 % BØUNDS ON LØWER PERCENTILES

| 1ST | 10TH | 20TH | 30TH | 40TH |
|---|---|---|---|---|
| 1.01864 | 2.25414 | 2.98328 | 3.61982 | 4.25222 |

LOWER 5 % BØUNDS ON UPPER PERCENTILES

| 60TH | 70TH | 80TH | 90TH | 99TH |
|---|---|---|---|---|
| 5.69495 | 6.62673 | 7.87346 | 9.88018 | 14.8946 |

UPPER 95 % BØUNDS ØN LØWER PERCENTILES

| 1ST | 10TH | 20TH | 30TH | 40TH |
|---|---|---|---|---|
| 2.06788 | 3.11737 | 3.91189 | 4.64786 | 5.40833 |

UPPER 95 % BØUNDS ØN UPPER PERCENTILES

| 60TH | 70TH | 80TH | 90TH | 99TH |
|---|---|---|---|---|
| 7.24331 | 8.50875 | 10.3243 | 13.6638 | 30.2364 |

MEAN ØF CØMMØN LØGS= .744276
SIGMA ØF CØMMØN LØGS= .250393
MEAN ØF NATURAL LØGS= 1.71376
SIGMA ØF NATURAL LØGS= .576552
50TH PERCENTILE= 5.54979

UPPER 95% LIMIT ØN INVERSE TRANSFØRM ØF MEAN ØF LØGS= 6.10191

MEAN ØF ØRIG. VARIABLE(ASSUMING IT IS LØG-NØRMAL)= 6.55328

SIGMA ØF ØRIG. VARIABLE(ASSUMING IT IS LØG-NØRMAL)= 4.11517