▼ Journal of Research (Science), Bahauddin Zakariya University, Multan, Pakistan.
Vol. 17, No. 2, April 2006, pp. 103-114 ISSN 1021-1012

DISCRIMINATION BETWEEN WEIBULL AND LOGNORMAL DISTRIBUTIONS FOR LIFETIME DATA

G. R. Pasha
Department of Statistics Bahauddin Zakariya University, Multan
email: drpasha@bzu.edu.pk
M. Shuaib Khan
Department of Statistics Govt Zamindar College, Gujrat
email: fahad_9_2@yahoo.com
Ahmed Hesham Pasha
Department of Electrical Engineering, Bahauddin Zakariya University, Multan
email: hesham01@gmail.com

Abstract
Reliability and failure data, both from life testing and from in-service records, are often modeled by the Weibull or Lognormal distributions so as to be able to interpolate and/or extrapolate results. The aim of this research is to discriminate between the Weibull and Lognormal distributions for complete samples. Both reliability models Weibull and Lognormal are then illustrated. Median rank regression (MRR) and maximum likelihood estimation (MLE) data- fitting methods are described and goodness-of-fit using maximum likelihood ratio (MLR) and most powerful invariant (MPI) tests. We find the Weibull distribution better for fitting to lifetime data when comparing with the Lognormal distribution.

Keywords: Maximum likelihood ratio test, median rank regression, most powerful invariant test, Weibull and Lognormal distributions.

INTRODUCTION
A life time distribution model can be any probability density function $f(t)$ defined over the range of time (0, ∞). The corresponding cumulative distribution $F(t)$ is very useful function as it gives probability that a randomly selected unit will fail by time t. Cohen [1951], Dubey [1966], Schlitzer [1966], Lock [1973], Gross and Lurie [1977], Bain [1978], Gibbons and Vance [1981], Lawless [1982] and Abernethy [1994] among many others find that there are a number of methods for fitting life data points to a distribution. We are going to discriminate two most frequently used life time distributions, Weibull and Lognormal distributions. This discrimination is based on goodness of fit tests after estimating the both lifetime models by two popular methods, namely, median rank regression (MRR) and maximum likelihood estimator (MLE). The goodness-of-fit test can test whether the complete life data

are from the underlying distribution or not. It is often found that the life data, which has been plotted on the relevant probability paper, fit both the Weibull and Lognormal lines very well. How do we objectively judge which model is the better choice? Abernethy [1996] and Fulton [1995] developed a graphical goodness-of-fit test and the p-value model.

Firstly, the life time distribution models: Weibull distribution and Lognormal distribution have been described. Secondly, the data description is given. The estimation method MRR, MLR and MPI tests are also explained in this section. Thirdly, the data analysis and discussion of the results is presented while last section concludes the research work.

THE LIFETIME DISTRIBUTION MODELS

In this section Weibull and Lognormal distributions have been described.

WEIBULL DISTRIBUTION

The Weibull probability distribution has three parameters η, β and t_0. It can be used to represent the failure probability density function (PDF) with time, so that:

$$f_w(t) = \frac{\beta}{\eta}\left(\frac{t-t_0}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-t_0}{\eta}\right)^{\beta}}; \quad \eta>0, \beta>0, t_0>0, -\infty<t_0<t \quad (1)$$

where β is the shape parameter (determining what the Weibull PDF looks like) and is positive and η is a scale parameter and is also positive, t_0 is a location or shift or threshold parameter (sometimes called a guarantee time, failure-free time or minimum life), t_0 be any real number, If $t_0 = 0$ then the Weibull distribution is said to be two-parameter.

LOGNORMAL DISTRIBUTION

The probability density function (PDF) for 3-parameter lognormal distribution is:

$$f_L(t) = \frac{\rho}{\sqrt{2\pi}(t-t_0)} \exp\left\{-\frac{\left[\ln\left(\frac{t-t_0}{\theta}\right)^{\rho}\right]^2}{2}\right\}, \quad \theta>0, \rho>0, t>0, -\infty<t_0<t, \quad (2)$$

where ρ is the shape parameter, θ is the scale parameter and t_0 is the location parameter. The units of ρ, θ and t_0 are the same as in the Weibull case. The Lognormal is said to be a two-parameter distribution when t_0 = 0. The restrictions on the values of t_0, θ, ρ for the Lognormal distribution are as stated in Eq. (1).

METHODOLOGY

We take the lifetime data of data 30 electric tubes lights used by Kale and Sinha [1971]. The data are given as Table A in Appendix. The *X* values in the data list represent actual electric tubes life data in hours. The data was analyzed by fitting Weibull and Lognormal models. The Median Rank Regression (MRR) and the Maximum Likelihood Estimation (MLE) were applied to estimate the unknown parameters. For the discrimination between the Weibull and the Lognormal, the MLR and MPI tests were used.

MEDIAN RANK REGRESSION (MRR) FOR COMPLETE SAMPLES

The lifetime data, when plotted on probability paper are approximately linear, so the parameters can be estimated by usual least squares (LS) method. Regressing *Y* on *X* minimizes the sum of squares of residual variation in the *Y* direction, whereas regressing *X* on *Y* minimizes it in the *X* direction. Berkson [1950] studied these two regressions and suggested that the scale with larger error should be regarded as the dependent variable. For life data analysis, the time-to failure, *X* always shows much more error than the median ranks especially for in-service failure data. So, Abernethy [1994] concluded that the better method for probability paper plot is to regress *X* on *Y*. He also showed that regressing *X* on *Y* has a better accuracy than regressing *Y* on *X*. So our study considers both regressions; *X* on *Y* and *Y* on *X*.

The Weibull cumulative distribution function (CDF), denoted by *F(t)*, is:

$$F_w(t) = 1 - e^{-(\frac{t-t_0}{\eta})^{\beta}} \tag{3}$$

The linear form of the resulting Weibull CDF can be represented by a rearranged version of Eq. (3):

$$\ln t = \frac{1}{\beta}\ln\ln\left(\frac{1}{1-F_W(t)}\right) + \ln\eta \tag{4}$$

Comparing this equation with the linear form $y = Bx + A$, leads to $y = \ln t$ and $x = \ln\ln\left\{ 1 \big/ 1 - F_W(t) \right\}$.

By minimizing β and η using the LS method one obtains:

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}, \tag{5}$$

and

$$\hat{\eta} = \exp\left(\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n\hat{\beta}}\right), \tag{6}$$

where n is the sample size and ˆ indicates an estimate. The mathematical expression for x_i and y_i are:

$$x_i = \ln\ln\left[\frac{1}{1 - F_W(t_i)}\right] \text{ and } y_i = \ln t_i .$$

$F(t_i)$ can be estimated by using Benard's formula, $\frac{i - 0.3}{n + 0.4}$, which is a good approximation to the median rank estimator [Tobias and Trindade 1986, Abernethy 1994]. The Benard's median rank was used because it showed the best performance and is the most widely used rank to estimate $F(t_i)$. The procedure for ranking complete data is as follows:

1. List the time to failure data from small to large.
2. Use Benard's formula to assign median ranks to each failure.
3. Estimate the β and η by Eqs. (5) and (6).

The $F(t_i)$ is estimated from the median ranks. Once $\hat{a}$ and $\hat{b}$ are obtained, the $\hat{\beta}$ and $\hat{\eta}$ can easily be obtained.

The estimator of ρ is the sample correlation coefficient, $\hat{\rho}$ given by:

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x - \bar{x})^2 \sum_{i=1}^{n}(y - \bar{y})^2}} \tag{7}$$

The Lognormal CDF is the integral of the PDF from 0 to time- to-failure *t*. It can be written in terms of the standard Normal CDF as:

$$F_L(t) = \Phi\left[\ln\left(\frac{t - t_0}{\theta}\right)^{\rho}\right] \tag{8}$$

The Lognormal CDF, when plotted against appropriate probability axes, appears linear and so can be represented by a rearranged version of Eq. (8) as:

$$\ln t = \frac{Z_p}{\rho} + \ln\theta \tag{9}$$

Comparing this with the linear form $y = B\,x + A$, leads to $y = \ln t$ and $x = Z_p$.

The same LS procedure, as applied for the Weibull distribution, yields.

$$\hat{\rho} = \frac{\sum_{i=1}^{n} x^2{}_i - \left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i} \tag{10}$$

and $$\hat{\theta} = \exp\left[\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n\hat{\rho}}\right] \tag{11}$$

where $x_i = z_p$, $y_i = \ln t_i$ and $z_p = \phi^{-1}(z_p)$ is the percentile of the standard Normal CDF, which is widely tabulated. Again $F(t_i) = \phi(z_p)$ can be estimated using Bernard's formula. The same ranking procedure was used as applied for the Weibull distribution. The rank table for Weibull is the same as that of Lognormal as shown in Table A.

THE MAXIMUM LIKELIHOOD RATIO (MLR) TEST

For MLR, the hypothesis setting is as follows:

H_0 : underlying distribution is the Weibull distribution.

H_1 : underlying distribution is the Lognormal distribution.

Level of significance is set at $\alpha = 0.01, 0.05, 0.10, 0.20$

The MLR test statistics (TS) is:

$$TS_{MLR} = \frac{1}{(\sqrt{2\pi\hat{\sigma}^2 e})\sqrt[n]{\prod_{i=1}^{n} t_i f_w(t_i)}} \tag{12}$$

Dumonceaux *et al.* [1973] also proposed the reverse hypothesis, as follows:

H_0 : Underlying distribution is the Lognormal distribution.

H_1 : Underlying distribution is the Weibull distribution.

Then, the test statistics (TS) of MLR becomes:

$$TS_{MLR} = \sqrt{2\pi\hat{\sigma}e}\sqrt[n]{\prod_{i=1}^{n} t_i f_w(t_i)} \tag{13}$$

The null hypothesis was rejected in the favor of alternative hypothesis whenever $TS_{MLR} \geq TS_{MLR}$ (tabulated).

THE MOST POWERFUL INVARIANT (MPI) TEST

Kent [1979] defined the most powerful invariant (MPI) test statistics for selection between the Weibull and Lognormal distributions given by:

$$TS_{MPI} = \frac{\hat{\beta}^{n-1}\Gamma(n)\left(\prod_{i=1}^{n} t_i\right)^{\hat{\beta}} (\hat{\sigma}\sqrt{2\pi})^{n-1}\sqrt{n}e^{\left(\frac{n-1}{2}\right)}}{\left(\sum_{i=1}^{n} t_i\hat{\beta}\right)^{n}} \tag{14}$$

where n is sample size, $\hat{\beta}$ and $\hat{\sigma}$ can be determined by the method of MLE.

If TS > 1, then the data could come from the Weibull distribution alternately,
If TS < 1, then the data could come from the Lognormal distribution.
So Eq. (12) can be rearranged as:

$$\frac{\left(\prod_{i=1}^{n} t_i\right)^{\hat{\beta}} e^{\left(\frac{n-1}{2}\right)}}{\left(\sum_{i=1}^{n} t_i \hat{\beta}\right)^{n}} \gtrless \frac{1}{\hat{\beta}^{\,n-1}\Gamma(n)(\hat{\sigma}\sqrt{2\pi})^{n-1}\sqrt{n}} \qquad (15)$$

After taking logarithms of both sides:

$$\hat{\beta}\left(\sum_{i=1}^{n}\ln t_i\right)+\frac{n-1}{2}-\left[n\ln\left(\sum_{i=1}^{n} t_i\hat{\beta}\right)\right] \gtrless -(n-1)\ln(\hat{\beta}\hat{\sigma}\sqrt{2\pi})-\ln[\Gamma(n)]-\frac{\ln(n)}{2} \qquad (16)$$

Let the left-hand side of the expression equate to A and the right-hand side of the expression equate to B, to simplify the inequality. Hence, if A > B, then this indicates that the samples could come from the Weibull distribution. Conversely, if B>A, then it indicates that the samples could come from the Lognormal distribution.

RESULTS AND DISCUSSION

The rank table is shown in Table A. The Weibull probability plot is obtained using Weibull Y-Bath™ as shown in Fig. 1. From Fig. 1, estimates of $\hat{\eta}$ = 656.5785 and $\hat{\beta}$ = 1.0435 were obtained.

By plotting the data (x on y) on probability paper the trend may appear curved, either concave down or up. These curvatures may indicate that the origin of the data is not the same as the zero from which the life data has been measured. In such cases a location parameter t_0 may be needed to make the origins coincide. Concave downward implies that a value $t_0 > 0$ is needed to convert the data. Concave upward implies that a value $t_0 < 0$ is needed. In practice concave downward situations are more often seen than concave upward [Abernethy 1994]. The case of t_0 >0 indicates that failures cannot occur until after a certain period of time has elapsed. During this period the units cannot fail. This is why t_0 is, sometimes, called guarantee parameter. The case of t_0 < 0 indicates that some duty may have occurred.

Fig. 2 illustrates how to deal with the 3-parameter Weibull distribution. Using Weibull *Y*-Bath™ $\hat{\eta}$ = 695.5714, $\hat{\beta}$ = 1.2555 and $\hat{t}_0$ = -31.6771 can be readily obtained. Fig. 2 shows the tube light failures using MRR. Note that the final result of η must be adjusted for t_0 to return to the original life scale, so that $\hat{\eta}$ = 695.5714 – 31.6771 = 663.8943. The result is $\hat{t}_0$ = -31.6771 when correlation coefficient CC = 0.9706 reaches the maximum using a graphical method.

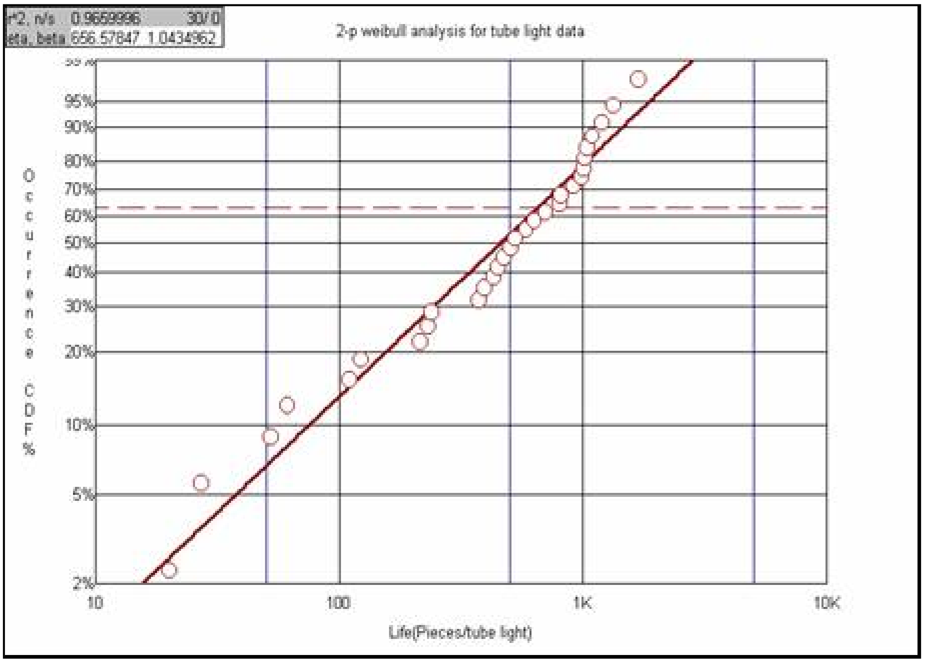

Fig. 1: 2-P Weibull Analysis for Tube Light Data.

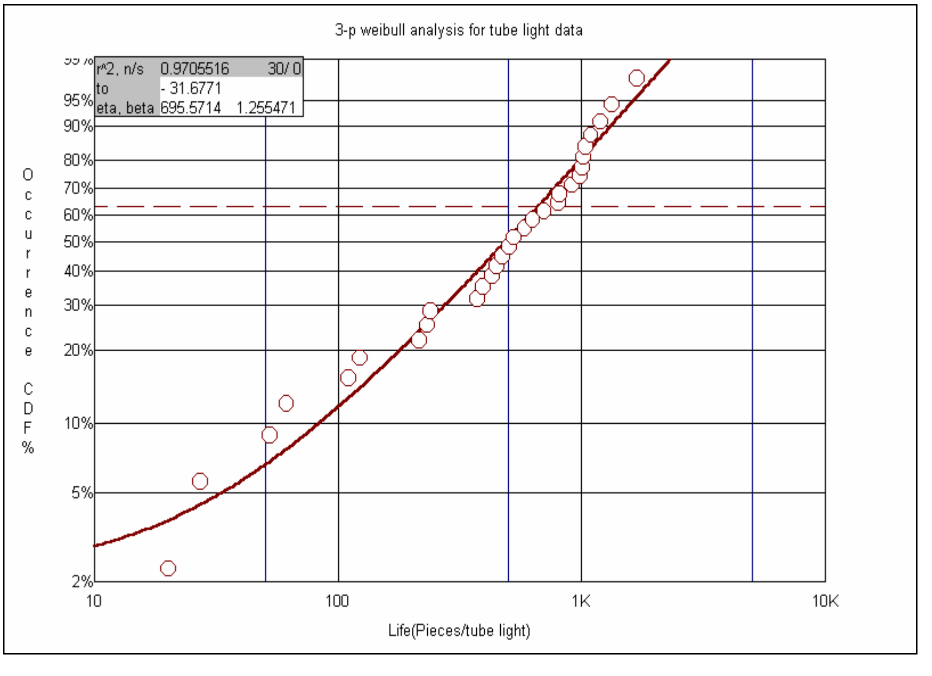

Fig. 2: 3-P Weibull Analysis for Tube Light Data.

When the data (*y* on *x*) was plotted on probability paper the trend appeared curved. Here again the tube light data [Kale and Sinha 1971] is taken to illustrate how to deal with the 2-parameter Weibull distribution. Using Weibull Y - Bath™ $\hat{\eta}$ = 668.9425, $\hat{\beta}$ = 1.0080 and similarly for 3-parameter Weibull distribution using Weibull Y-Bath™ $\hat{\eta}$ = 705.3958, $\hat{\beta}$ = 1.2201, $\hat{t}_0$ = -31.6221 can be readily obtained. Results show the tube light failures using MRR. Note that the final result of η must be adjusted for to t_0 return to the original life scale, so that $\hat{\eta}$ = 705.3958 – 31.6221 = 673.7737. The result is $\hat{t}_0$ = -31.6221 when *CC* = 0.9706 reaches the maximum using a graphical method.
Win SMTTH™, as shown in Fig. 3, this terminology stands for anti-log of the log-value mean. The sdF is a multiplier / divisor and represents the anti-log of life log-value standard deviation [Fulton 1995]. The estimated values obtained were muAL = 385.4 = $\hat{\theta}$ and sdF = 3.176. The value of sdF = 3.176 leads to a result for $\hat{\rho}$ = 0.8654.

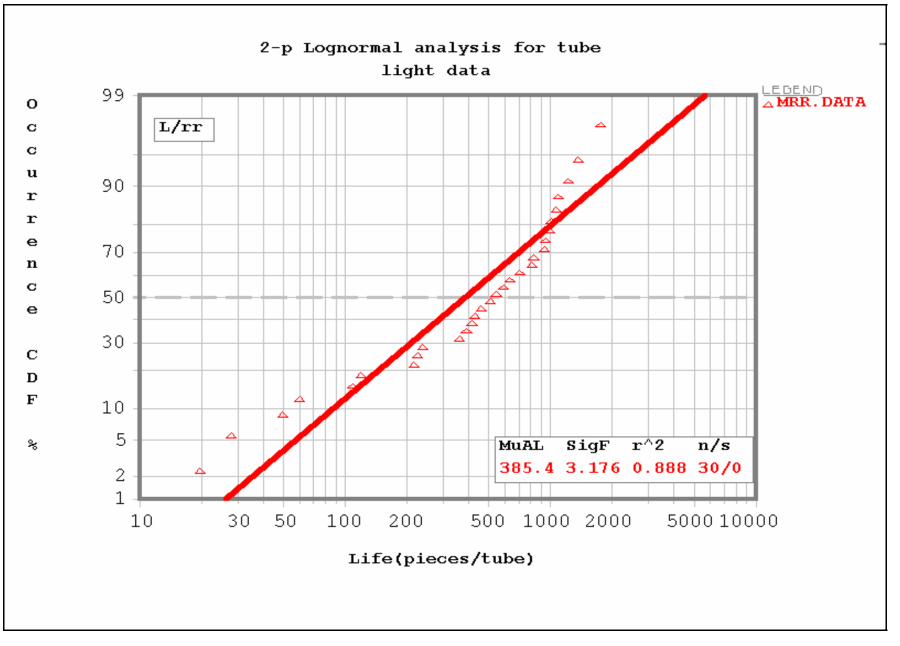

Fig. 3. 2-P Lognormal Analysis for Tube Light Data.

The 3-parameter Lognormal analysis for the tube light data are similar to the Weibull case as shown in Fig. 4. The estimated values obtained were muAL = 1256 = $\hat{\theta}$, $\hat{t}_0$ = -723.2 and sdF = 1.411. The value of sdF = 1.411 leads to a result for $\hat{\rho}$ = 2.9045. The final result of θ must be adjusted for t_0 to return to the original life scale. So $\hat{\theta}$ = 1256.0-723.2 = 532.8.

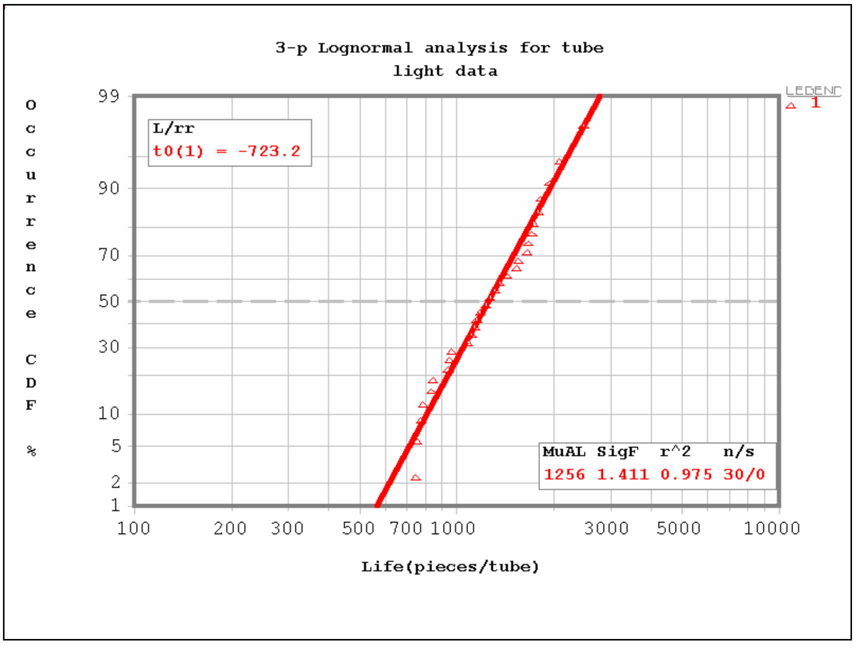

Fig. 4: 3-P Lognormal Analysis for Tube Light Data.

The results of the MRR method are summarized as presented in Table 1.

Table 1: Summary of Results for Tube Light Data Using MRR Method (n = 30).

Distributions	Weibull			Lognormal		
Method	Parameters	2-P	3-P	Parameters	2-P	3-P
MRR	η	656.5785	695.5714	θ	385.400	1256
	β	1.0435	1.2555	ρ	3.176	1.411
	t_0	0.0	-31.6177	t_0	0.0	-723.2
	CC	0.9660	0.9706	*CC*	0.888	0.975

GOODNESS-OF-FIT

There are many methods of measuring goodness-of-fit such as Chi-square test and MLR test etc. Dumonceaux *et al.* [1973], Lawless [1982], O'Connor [1991] and Dodson [1994] used other methods such as MPI tests to evaluate the difference between the Weibull and Lognormal distributions. The MLR and MPI tests were considered for the present study.

Using estimated values of unknown parameters in Eq. (12), the value of MLR statistics was found to be $TS_{MLR} = 0.8738$. When this value was compared with the critical values given in Table 2 for different α, the null hypothesis H_0 was accepted in all the cases, which led to the conclusion that the Weibull is a statistically significantly better fit than the Lognormal.

By considering reverse hypothesis proposed by Dumonceaux *et el*. [1973], one has $TS_{MLR} = 1.1444$. The same conclusion can again be drawn for the better fit of Weibull distribution over that of Lognormal.
Finally for MPI test, A = -101.569 > B = -105.0.61. The results of the MPI tests were also found to be consistent with MLR tests. Therefore, It can be deduced that the data could have come from the Weibull distribution.

Table 2: Critical Values of MLR Test for Discriminating between the Weibull and Lognormal Distributions.

n	$\alpha = 0.2$		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
	TSW_{MLR}	TSL_{MLR}	TSW_{MLR}	TSL_{MLR}	TSW_{MLR}	TSL_{MLR}	TSW_{MLR}	TSL_{MLR}
20	1.008	1.015	1.041	1.038	1.067	1.028	1.120	1.144
30	0.991	0.993	1.019	1.020	1.041	1.044	1.088	1.095
40	0.980	0.984	1.005	1.007	1.026	1.028	1.063	1.070
50	0.974	0.976	0.995	0.998	1.016	1.014	1.045	1.054

CONCLUSIONS

The Weibull distribution and Lognormal distributions are extensively used in reliability and life testing. We have concluded that while comparing these two distributions, Weibull distribution provides better fit to lifetime data. We also note that MLR and MPI tests are consistent with this discrimination between the Weibull with the Lognormal distribution.

References

Abernathy, R.B. (**1994**) “The New Weibull Handbook”, SAE Professional Development.

Abernathy, R.B. (**1996**) “The New Weibull Handbook”*,* 2nd ed., Dept. AT Houston, Texas, USA..

Abernathy, R.B. and Fulton, W. (**1995**) Weibull NEWS*.* Fall.

Bain, L..J. (**1978**) “Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods”*,* New York, USA..

Berkson, J. (**1950**) “Two Regressions”, *J. Amer. Stat. Assoc*., **45**, 164-180.

Cohen, A.C. (**1951**) “Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood”, *J. Amer. Stat. Assoc.*, **46**, 206-212.

Dodson, B. (**1994**) “Weibull Analysis”, ASQC Quality Press, Milwaukee, Wisconsin, USA..

Dubey, S.D. (**1966**) “Comparative Performance of Several Estimators of the Weibull Parameters”, *Proceedings of the 20th Technical Conference of the American Society for Quality Control*, 723-735.

Dumonceaux, R., Antle, C.E. and Haas, A. (**1973**) “Likelihood Ratio Test for Discrimination between Two Models with Unknown Location and Sale Parameters”, *Technometrics*, **15**, 19-27.

Fulton, W. (**1995**) "Fulton Findings, Weibull SMITH™, Monte Carlo SMITH™ & Visual SMITH™ software", 1251 W. Sepulveda Blvd. #800, Torrance, CA 90502, USA.

Gibbons, D.I. snd Vance, L.C. (**1981**) "A Simulation Study of Estimators for the 2-Parameter Weibull Distribution", *IEEE Transactions on Reliability*, R-**30**(1), 61-66.

Gross, A.J. and Lurie, D. (**1977**) "Monte Carlo Comparisons of Parameter Estimators of the 2-parameter Weibull Distribution", *IEEE Transactions on reliability*, R-**26**(.5), 356-358.

Kale, B.K. and Sinha, S.K. (**1971**) "Estimation of Expected Life in the Electric Components", *Technometrics*, **14**, 137-144

Kent, J.E.A. (**1979**) "An Efficient Procedure for Selecting among Five Reliability Models", North Carolina State University, Raleigh, USA.

Lawless, J.F. (**1982**) "Statistical Models and Methods of Lifetime Data", John Wiley, New York, USA.

Lock, M.O. (**1973**) "Reliability, Maintainability, and Availability Assessment", Hayden Book Co. New Jersey, USA.

O'Connor, P.D.T. (**1991**) "Practical Reliability Engineering", 3rd ed., John Wiley, New York, USA.

Schlitzer, L.D. (**1966**) "Monte Carlo Evaluation and Development of Weibull Analysis Techniques for Small Sample Sizes", May 2-5.

Tobias, P.A. and Trindade, D.C. (**1986**) "Applied Reliability", Van Nostrand Reinhold Co. New York, USA.

Appendix

Table A: Weibull Analysis for Electric Tubes Light data.

Rank Table				
Set	Point	*x*-value	Quantity	Benard's Rank
1	01	20.15414	1	2.303E-02
1	02	26.61217	1	5.592E-02
1	03	53.55068	1	8.882E-02
1	04	61.43307	1	0.1217105
1	05	111.5966	1	0.1546053
1	06	118.1282	1	0.1875000
1	07	213.5254	1	0.2203947
1	08	227.5179	1	0.2532895
1	09	234.7571	1	0.2861842
1	10	370.9544	1	0.3190790
1	11	387.6415	1	0.3519737
1	12	421.5968	1	0.3848684
1	13	427.3025	1	0.4177631
1	14	472.4213	1	0.4506579
1	15	515.6355	1	0.4835526
1	16	523.4068	1	0.5164474
1	17	554.0378	1	0.5493421
1	18	595.6667	1	0.5822368
1	19	718.2560	1	0.6151316
1	20	812.9323	1	0.6480263
1	21	816.7766	1	0.6809211
1	22	936.1332	1	0.7138158
1	23	952.9595	1	0.7467105
1	24	984.0899	1	0.7796053
1	25	992.5800	1	0.8125000
1	26	1041.677	1	0.8453947
1	27	1080.865	1	0.8782895
1	28	1143.756	1	0.9111842
1	29	1270.303	1	0.9440789
1	30	1627.440	1	0.9769737