ISSN 1684 – 8403
Journal of Statistics
Vol: 13, No.1 (2006)

Empirical Analysis of The Weibull Distribution for Failure Data

G.R Pasha[1], M. Shuaib Khan[2], Ahmed Hesham Pasha[3]

Abstract

In this paper we present the comprehensive analysis for complete failure data. The aim of this research is the *Empirical Analysis of the Weibull Distribution for Failure Data.* We access the Weibull distribution assumptions of a data set. Median rank regression (MRR) for data- fitting method is described and goodness-of-fit using correlation coefficient is applied. We use the simulation technique to present confidence bound.

Key Words

Weibull distribution, Hard disk failure data, Electronic Failures, goodness-of-fit tests, confidence bound.

1. Introduction

A life time distribution model can be any probability density function $f(t)$ defined over the range of time from $t = 0$ to $t =$ infinity. The corresponding cumulative distribution $F(t)$ is very useful function, as it gives probability that a randomly selected unit will fail by time t. Abernethy (1994) suggested a number of methods for fitting the life time data points to a distribution. We are going to investigate one of the most frequently used method, median rank regression (MRR), which is popular in industry. After we have

[1] Department of statistics Bahauddin Zakariya University Multan.
e-mail: 'drpasha@bzu.edu.pk'
[2] Department of statistics The Islamia University of Bahawalpur
e-mail: 'skn_801@yahoo.com'
[3] Department of Electrical Engineering Bahauddin Zakariya University Multan.
e-mail: 'Hesham01@gmail.com'
Received Dec. 2006, Accepted May, 2007

fitted the distribution to the data points we can then compute the parameters of that distribution using the MRR method. We can use the MRR method to solve for the parameters. The goodness-of-fit test can test whether the complete life data are from the weibull distribution or not. Abernethy and Fulton (1995, 96) idea are developed a graphical goodness-of-fit test. For a graphical goodness-of-fit test model uses the correlation coefficient (CC), which for the 2-parameter Weibull distribution, has been calculated. The resulting CC values can be used not only to compare fits by MRR methods but also to indicate how good a particular fit is. The primary advantage of Weibull analysis is the ability to provide reasonably accurate failure analysis and failure forecasts with extremely small samples Chi-chao, (1997). Solutions are possible at the earliest indications of a problem without having to "crash a few more". Another advantage of weibull analysis is that it provides a simple and useful graphical plot. In this paper the data plots are extremely important to the engineers.

1.1 Theoretical Background

1.1.1 Weibull Distribution

The Weibull probability distribution has three parameters η, β and t_0. Aberntthy 1994 express the shape and scale parameter of the Weibull distribution by $W(\eta, \beta)$. It can be used to represent the failure probability density function (PDF) with time, so it is defined as:

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

Where β is the shape parameter (determining what the Weibull PDF looks like) and is positive and η is a scale parameter (representing the characteristic life at which 63.2% of the population can be expected to have failed) 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 can be any real number, If $t_0 = 0$ then the Weibull distribution is said to be two-parameter Chi-chao, (1997). When β = 1, the distribution is the same as the exponential distribution for the density function. When $\beta = 2$, it is known as the Rayleigh distribution for the density function. When $\beta = 2.5$, then the shape of the density function is similar to the Lognormal shape of function. When β = 3.4 is then the shape of the density function is similar to the Normal shape of function

2. Method of Analysis

2.1 Description of data

Computer Hard Disk Drive Failures data is taken to illustrate how to use the MRR methods, For the *Empirical Analysis of the Weibull Distribution for Failure Data.* The data recorded the number of failures occurring within fixed time. The t values in the data list represent actual time to failures life data in hours. There are 16 Computer Hard Disk Drive Failures life data Computer Hard Disk Drive Failures "Source: Dev Raheja, Assurance Technologies".

2.2 Weibull Analysis

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

$$F_w(t) = 1 - e^{-(\frac{t-t_0}{\eta})^{\beta}} \qquad (2)$$

The liner form of the resulting Weibull CDF can be represented by a rearranged version of equation (2):

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

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

If we minimize β and η by using the LS method then we obtain:

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}, \quad (3)$$

and
$$\hat{\eta} = \exp\left(\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n\hat{\beta}}\right), \quad (4)$$

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 (Abernethy 1994 , Chi-chao, 1997, Tobias and Trindade 1986). We use the Benard’s median rank because it shows the best performance and it is the most widely used 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 equations (3) and (4).

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

3. Analysis Output and Result

Computer Hard Disk Drive Failures data is taken to illustrate how to use the MRR method. The Weibull method converting the recorded times into data points for linear regression using the above mention weibull analysis we perform linear regressions for the data points. From the linear regression of weibull analysis we estimate β and η. There are 16 Computer Hard Disk Drive Failures life data shown as follows

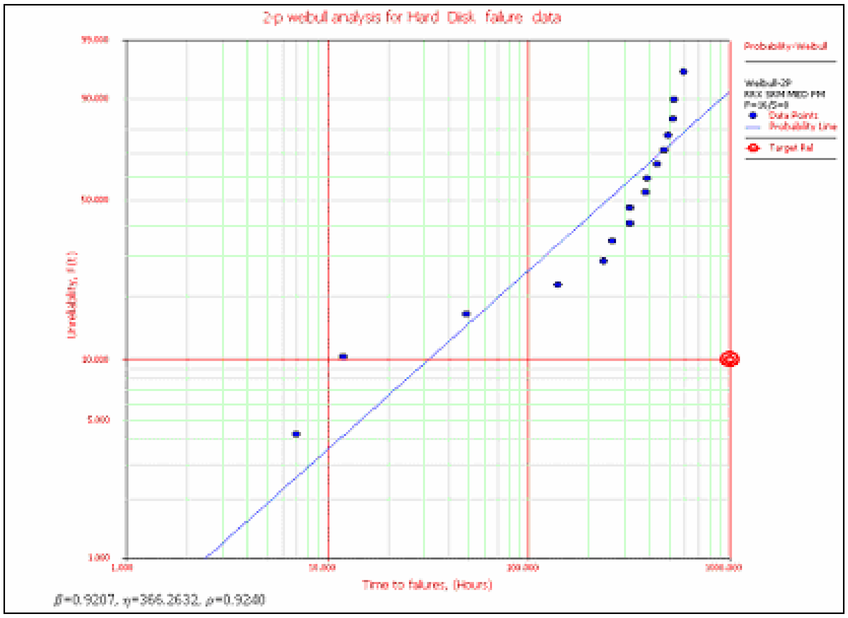

Fig.1 2-P Weibull Analysis for Hard Disk failure Data

Using Weibull++6 $\hat{\eta}$ = 366.2632, $\hat{\beta}$ = 0.9207 and ρ = 0.9240 can be readily obtained. Fig. 1 shows the Hard disk drive failures using MRR. The horizontal scale is a measure of failures life. The vertical scale is the cumulative percentage failed .The two defining parameters of the weibull line are the slope, beta, and the characteristic life, eta. The slope of the line β is particularly significant and may provide a clue to the physics of the failure. We have age of the parts that are falling. Here age is the operating time. Here Beta <1 Implies Infant Mortality. Here the failure modes for Hard disk drive are beta<1, and the Hard disk drive survives infant mortality, it will improve with age as the hazard rate declines with age. When the shape parameter is less than one it conclude that the life span of these drives is less than one year. Conditional on survival the hazard rate decreases and the reliability increases. Overhauling such a Hard disk drive is not appropriate, as old parts are better than new.

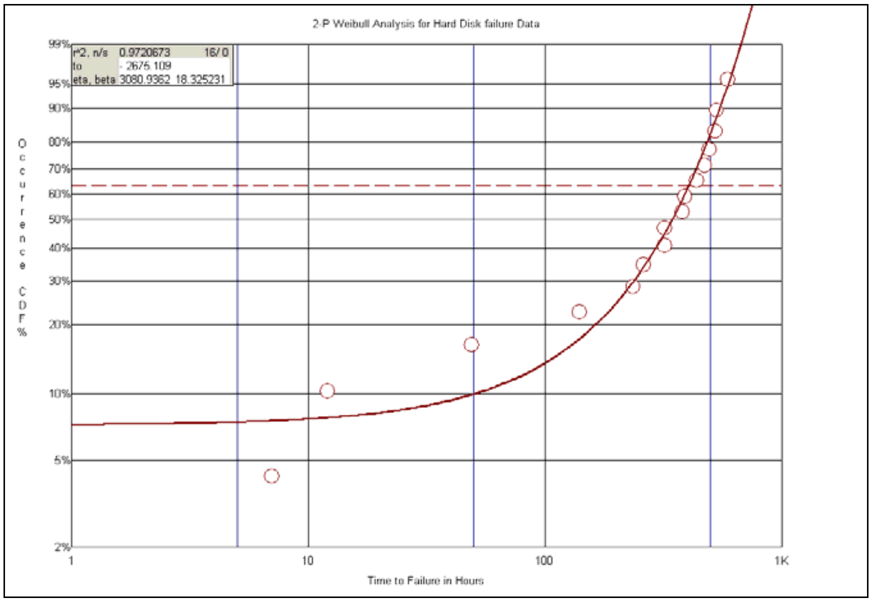

Fig.2 3-P Weibull Analysis for Hard Disk failure Data

Using Weibull Y-Bath™ $\hat{\eta}$ = 3080.9362 $\hat{\beta}$ = 18.325231 and $\hat{t}_0$ = -2675.109 can be readily obtained. Fig. 2 shows the hard disk failures using MRR. Note that the final result of η must be adjusted for to t_0 return to the original life scale, so $\hat{\eta}$ = 3080.9362 -2675.109 = 405.8272. The result is $\hat{t}_0$ = -2675.109 when cc = 0.98594 reaches the maximum and the critical correlation coefficient ccc = 0.94739 Using a graphical method. In fig.2 the pattern of failure are closed to the estimated line. From this pattern it concludes that three parameter weibull distribution follow the hard disk failure data. Here $\hat{t}_0 < 0$ it shows that concave is upward. From the linear regression of weibull analysis we estimate β and η. The slope and the y-intercept for the regression line of figure 1 are 0.9207and 366.2632 respectively. We then obtain the weibull parameters $\hat{\beta}$ = 0.9207, $\hat{\eta}$ = 366.2632 and the correlation coefficient ρ = 0.9240. The weibull probability plot is obtained using Weibull++6 as shown in the following Figures from 1 and 3-6. Here the horizontal scale is Hard disk Failure data in hour's x .In figures (1) the vertical scale is the cumulative density function (CDF), the proportion of the units that will fail up to age x in percent. The Statistical symbol for the CDF is $F(x)$, the probability of failure up to time x. Here in this graph the doted points which show the failure components. In figures (3) the vertical scale is the Value. In figures (4) the vertical scale is the failure rate (FR), The hazard function (HF) (also known as instantaneous failure rate), denoted by h(*t*), is defined as *f(t)/R(t)*. The units for *h*(*t*) are probability of failure per unit of time, in hours. Here $\beta < 1$, the hazard function is continually decreasing which represents Infant Mortality. In figures (5) the horizontal scale is hard disk failure data in hours x. The vertical scale is the failure time line. In this graph the star line shows the failure component. In figures (6) the horizontal scale is eta the characteristic life (η) and the vertical scale is beta (β) for the Contour Analysis for hard disk failure data. Here in these graph we find (99, 95, 90, 85, 75) ($1-\alpha$) percent confidence interval is the range of values, bounded above and below, within which the true value is expected to fall. It measures the statistical precision of our estimate. The probability that the true value lies within the interval is either zero or one, If we use 90% or above confidence interval then it will contain the true value for reliability intervals. Here the likelihood contour plots for the parameter eta and beta do not intersect; this would indicate significant differences with confidence levels. We have age of the hard disk failure data; here age is the operating time. Here Beta <1 Implies Infant Mortality. The Infant Mortality shows the decreasing hazard function. In decreasing hazard function new item in this life period has a larger probability of failing than an old item. This shows that the Hard disk may initially have high failure rates. The infant mortality stems from the high mortality of infants. This shows that the drive may initially have high failure rates. Therefore, betas less than one lead us to suspect:

- **Inadequate burn-in or stress screening,**
- **Production problems, misassemble, quality control,**
- **Overhaul problems ,**
- **Solid state electronic failures**,

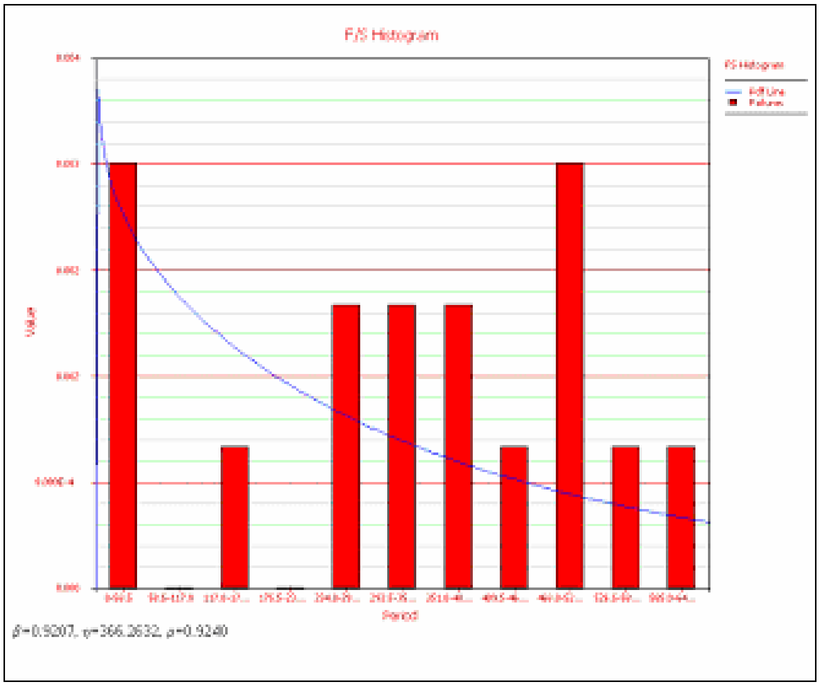

Fig.3 2-p Weibull Histogram Analysis for Hard disk Failure data

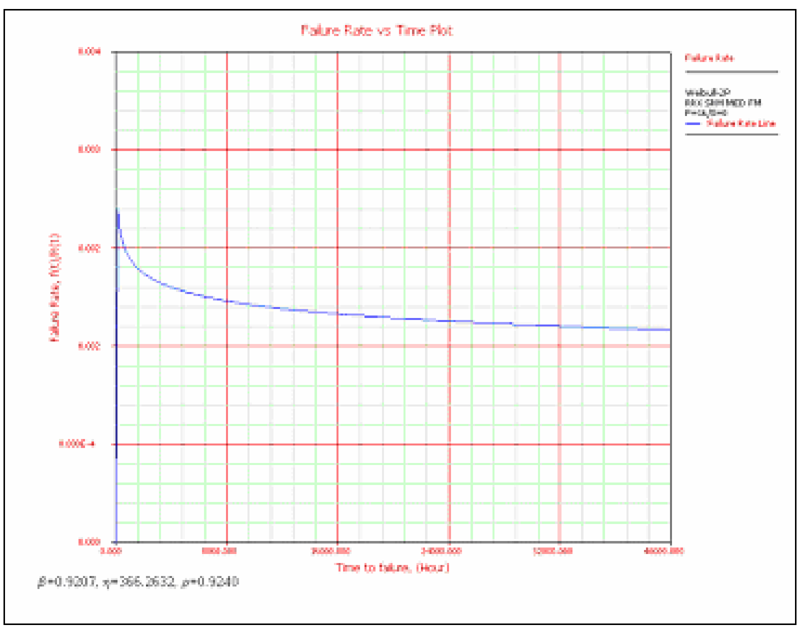

Fig.4 2-p Weibull Failure Rate Analysis for Hard disk Failure data

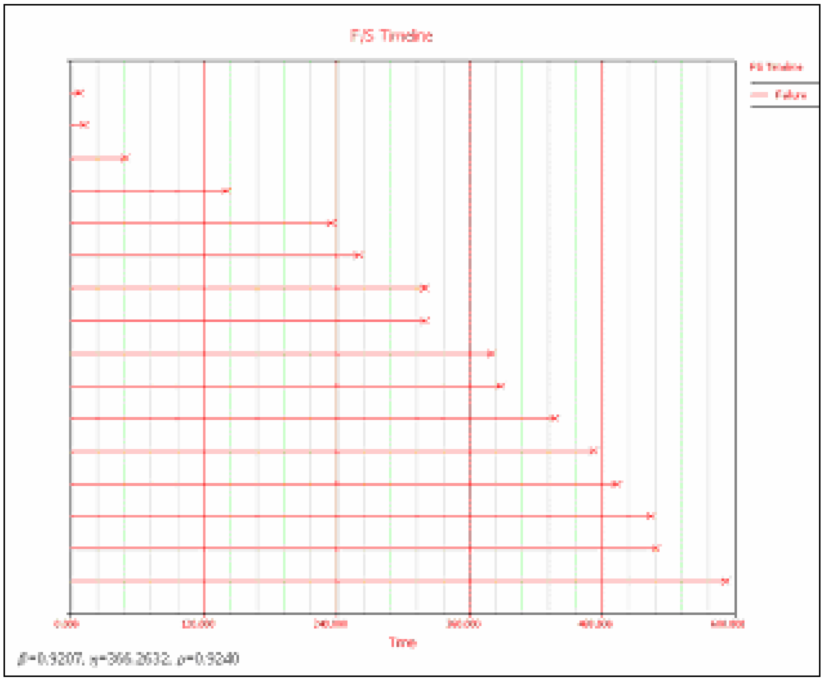

Fig.5 2-p Weibull Time Analysis for Engine fan life data

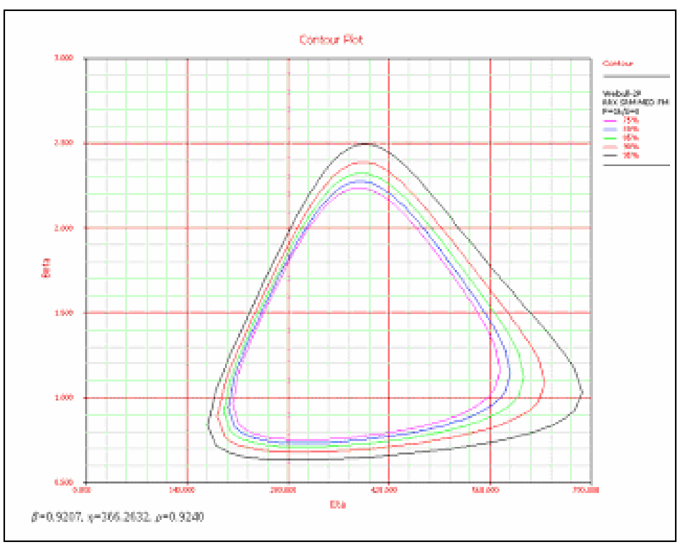

Fig.6 2-p Weibull Contour Analysis for Hard disk Failure data

The results of the sample about hard disk failure data using the MRR method are summarized in Table 1.

Table 1. **Summary of results for hard disk failure data using MMR Method (n = 16)**

Distributions		Weibull Distribution	
Method	Parameters	2-P	3-P
	η	366.2632	3080.9362
MRR	β	0.9207	18.325231
	t_0	0	-2675.109
	Cc	0.9240	0.98594

As we can see, the 3-parameter Weibull model CC = 0.98594 show the curvature. But the 2-parameter Weibull plot do not show the curvature see Figure 1. Using the 2-parameter models, considering the 2-parameter Weibull CC = 0.9240. It follows that 3-p weibull gives consistent result than 2-p weibull distribution.

3.1 Goodness-of-fit

In statistics, there are many methods of measuring goodness-of-fit such as Anderson darling and kolmogorov simernov tests are applied but we prefer the simple correlation coefficient. It is ideal for testing the goodness of fit to a straight line. The correlation coefficient "r" is intended to measure the strength of a linear relationship between two variables. As Weibull always have positive slopes, they will always have positive correlation coefficients. The closer "r" is to one the better the fit. The distribution of the correlation coefficient from ideal weibulls based on median rank plotting positions, For the 2-P Weibull Analysis for Hard Disk failure Data Here coefficient of determination $r^2 = 0.85377 <$ ccc =0.89169 so the fit is bad. While the 3-P Weibull Analysis for Hard Disk failure Data the coefficient of determination $r^2 = 0.97207 >$ ccc =0.94739 so the fit is good.

4. Conclusions

The Weibull is extensively used in reliability and life testing. The weibull distribution is fitted the life data very well. We have concluded that while comparing this distribution for the two and three parameter, from the comparisons of the above results here the three parameter weibull distribution provides better results than the two parameter Weibull distribution. But we also note that from the correlation coefficients the three parameter weibull distribution is consistent for the hard disk failure data.

References:

1. Abernathy, R. B. (1994). The New Weibull Handbook. SAE Professional Development.

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

Abernathy, R. B. (2004). The New Weibull Handbook, *4th* Edition, Dept. AT Houston, Texas 77252-2608, USA.

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

4. Berkson, J. (1950). Two Regressions Journal of the American Statistical Association. Vol. 45, (164-180).

5. Lawless, J. F. (1982). Statistical Models and Methods of life time data, J WILEY, New York, USA.

6. Liu, Chi-chao, (1997). A Comparison between the Weibull and Lognormal Models used to Analyze Reliability Data. P.hd thesis University of Nottingham, UK

7. Schlitzer, L D. (1966) Monte Carlo Evaluation and Development of Weibull Analysis Techniques for Small Sample Sizes.

8. Source of Computer Hard Disk Drive Failures: Dev Raheja, Assurance Technologies.

APPENDIX

Table A :Analysis of Hard Disk Failure Data

Time in hr	P_X
7	0.042683
12	0.103659
49	0.164634
140	0.22561
235	0.286585
260	0.347561
320	0.408537
320	0.469512
380	0.530488
388	0.591463
437	0.652439
472	0.713415
493	0.77439
524	0.835366
529	0.896341
592	0.957317

For Two Parameter Weibull Distribution MRR

Table B: 90% C.I for MRR in Monte Carlo Analysis

RUN	r^2	ETA	BETA
Target	0.8537700	366.2632	0.9207271
Lower	0.8598219	209.6378	0.5971654
Upper	0.9944915	563.9925	1.4068030

Table C: 95% C.I for MRR in Monte Carlo Analysis

RUN	r^2	ETA	BETA
Target	0.8537700	366.2632	0.9207271
Lower	0.8826149	180.4981	0.5927805
Upper	0.9864513	634.3573	1.7665380

Table D: 99% C.I for MRR in Monte Carlo Analysis

RUN	r^2	ETA	BETA
Target	0.8535761	366.5067	0.9178682
Lower	0.8373939	177.2180	0.5272902
Upper	0.9780507	566.5307	1.6812440

For Three Parameter Weibull Distribution MRR

Table E: 90% C.I for MRR in Monte Carlo Analysis

RUN	r^2	To	ETA	BETA
Target	0.9720674	-2675.109	3080.936	18.32523
Lower	-3.27274E+15	-45312.96	5.033961E-05	6.51398E-3
Upper	1.11142E+15	776.7932	7.462147E+17	328.3123

Table F: 95% C.I for MRR in Monte Carlo Analysis

RUN	r^2	To	ETA	BETA
Target	0.9720674	-2675.109	3080.936	18.32523
Lower	-1.04656E+15	-42812.13	281.3321	6.4677E-03
Upper	3.132236E+14	746.8779	8.249263E+17	318.9428

Table G :99% C.I for MRR in Monte Carlo Analysis

RUN	r^2	To	ETA	BETA
Target	0.9720674	-2675.109	3080.936	18.32523
Lower	-1.93351E+15	-1542434	176.3847	6.67443E-3
Upper	1.454115E+14	657.758	2.813135E+17	10744.25