PUBLICATIONS
ANALYSIS OF PRODUCTION DOWNTIMES IN THERMAL POWER PLANT USING EXPONENTIAL FAILURE MODEL – A CASE STUDY

C.P.S. Hungund* and Shrikant Patil**
*Professor, Department of Statistics, Gulbarga University, Gulbarga
e-mail id: cpshgu@gmail.com
**Research Scholar, Department of Statistics, Gulbarga University, Gulbarga
e-mail id: shrikant07.inc@gmail.com

Received on 12 April 2013: Accepted on 30 May 2013

ABSTRACT

Now a day’s Karnataka State is witnessing electricity load shedding due to various reasons among which, technical problems play vital role. These technical problems lead to failure of equipments like Boiler, Furnace and so on, leading to the electricity load shedding. Smoother flow of electricity generation is very essential to reduce the electricity load shedding. For this it is required to understand and analyze the failure patterns or production downtimes in a systematic manner.

Thus in this paper, we have developed a model based on failure patterns of seven thermal power units located at Shaktinagar, Raichur., Karnataka. For the collected data, Exponential Model is fitted and tested for its Goodness of Fit. Conclusions were drawn based on the result obtained.

Keywords: Exponential Failure Model, Electricity load shedding, Production Downtimes (Repair Hours), Goodness of Fit.

1. INTRODUCTION

For any nation to achieve significant industrial and economic growth, power sector is very important. The poor development of power sector can have far-reaching consequences on the economic growth of any nation as every company/industry requires energy to produce goods. Electricity is the primary source of energy and the development of electricity sector is important for improving productivity and hence economic growth. Power is derived from various sources. These include

- Thermal Power,
- Hydropower or Hydroelectricity,
- Solar Power,
- Biogas Energy,
- Wind Power and so on.
C.P.S. Hungund and Shrikant Patil

In all the above sources of power generation, system reliability is the most important concern and generation of the power mainly depends on the effective functioning of the system. In order to estimate the reliability of a system one must concentrate on the failure patterns or downtime of the production units.

For the present study, we have chosen Raichur Thermal Power Station (RTPS) to study thermal units. In this thermal power generation a number of equipments and systems are involved which are listed as follows:

- Steam generator
- Boiler furnace and steam drum
- Superheater
- Reheater
- Steam turbine
- Auxiliary systems
- Fuel preparation system.

Due to continuous running of above equipments, there is wear and tear to the part of the equipment leading to their failure. In this study, a unit represents the system consisting of all the above mentioned equipments involving in electricity productions. There are such seven power generating units at RTPS. Data on Working Hours and Repair Hours are collected from the seven units during the period 2004-2011.

Though the thermal stations in general are highly reliable sources of power generation, many factors and parameters govern their reliability. Hence management of a thermal power station is a highly challenging job not only for technical reasons but also for strategic importance of the end product, i.e., electricity exerts a lot more concentration over the control and coordination between its various sub-systems.

In the present study we have used Exponential Distribution to fit the data through Chi-Square test of Goodness of Fit (GoF). The study can be extended to show that even Weibull, Normal or Lognormal distributions give a good fit for repair hours of thermal power plant. In the following sections we have discussed survey of literature, objectives of the study, methodology, fitting of Exponential Distribution for repair hours of seven power units of thermal power plant at Raichur, analysis and conclusion.

2. SURVEY OF LITERATURE

Krishna Reddy (2008) dealt a case of “Modeling the Causes of Production Downtimes: An Empirical Study of a Thermal Power Generating Unit”. In the study, the author has made an attempt to understand and minimize the causes of production downtimes in thermal power generation. The data on working hours and breakdown hours of power units are collected. It is essential to understand and model various causes of downtimes for improving the performance, which can lead to effective decision making. An attempt is made to find which distribution is best fitted for working time between successive failures (working hours) and downtime (repair hours). Exponential distribution is a commonly assumed model of failure time and repair time. However, even the system failure time or working time is composed
of additive components, each independently, approximately, and exponentially distributed. One cannot expect the observed failure and working times to follow the Exponential distribution; instead, one can expect the Weibull or a Log-normal distribution with non-constant failures rates to give a good fit. Chi-square test is used to test the GoF.

Romeu (2004) discussed GoF test using various distributions in "The Chi-Square: a Large Sample Goodness of Fit Test". In the study, the author has developed a model to fit the data using Normal Distribution, Exponential Distribution and Weibull Distribution. Then Chi-Square test was applied to know the underlying distribution supports the data or not. The author has taken several examples for Normal, Lognormal, Exponential and Weibull distributions and showed how the GoF can be dealt via Chi-Square test. The paper discusses importance of GoF assessments of statistical distributions. Also showed that, for the small sample GoF cannot be done through Chi-Square test. For the number of observations per cell is too small for the GoF test statistic to converge to its Chi-Square underlying distribution. In such cases, one can use Cumulative Distribution Function (CDF) based distance GoF tests, such as the Anderson-Darling and Kolmogorov-Smirnov test.

Kohichi Nakamura, (1981) University of Tokyo had worked on "Reliability analysis of thermal power units". In the paper he had obtained statistical estimation of operating reliability, failure rate and maintainability. It has been assumed that all the units have the same failure rate and the same repair rate. Weibull Distribution analysis was used to derive the:

- Failure rate function
- Operating reliability function
- Maintainability function and
- Repair rate function of thermal unit.

The Weibull Distribution analysis used for the following reasons:

- Distribution function forms of operating time and down time can be estimated.
- Operating reliability function and maintainability function of thermal units with different distribution function forms can be estimated statistically.
- Operating Parameter characterizing Weibull Distribution can be estimated easily using Weibull Probability.

The paper proves that Weibull Distribution can be used to estimate various aspects of reliability functions of thermal power units.

3. OBJECTIVES

The objectives of the present study are:

- To fit the exponential distribution for repair hours of seven power units of a thermal power plant.
- To test the assumption of exponentiality through the chi-square test of goodness of fit.
4. METHODOLOGY

Here an effort is made to fit the data using exponential distribution for
downtimes (repair hours). Exponential distribution is commonly assumed model of
time to repair hours. According to the procedure suggested by Romeu, (2004),
establish the failure distribution from the data, before one can correctly implement
the test procedures. We then estimate the distribution parameters (e.g., mean and
variance); such process yields the "composite" distribution hypothesis called the null
hypothesis. The negation of the assumed distribution is called the alternative
hypothesis. We then test the assumed distribution using the data set. Finally, H0 is
rejected whenever any one (or more) of the several elements in hypothesis H0 is not
supported by the data.

The GoF tests are the statistical procedures that allow establishing whether an
assumed distribution is correct. The Chi-Square test is conceptually based on the
probability density function (PDF) of the assumed distribution. If this distribution is
correct, its PDF should closely encompass the data range (of X). We thus select
convenient values in this data range that divide it into several subintervals. Then, we
compute the number of data points in each subinterval. These are called “observed”
values. Then, we compute the number that should have fallen in these same
subintervals, according to the PDF of the assumed distribution. These are called the
"expected" values and the Chi-Square test requires at least five of them in every
subinterval. Finally, we compare these two results. If they agree (probabilistically)
then the data supports the assumed distribution. If they do not, the assumption is
rejected. The formula (statistic) that uses the differences between “expected” and
“observed” values to test the GoF follows a Chi-Square distribution. Hence, the name
Chi-Square tests.

The distribution of test statistic is as follows:

\[ \chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \]  

(4.1)

where

- \( E_i \) is expected number of data points (\( E_i \geq 5 \))
- \( O_i \) is observed number of data points
- \( k \) is total number of cell or subintervals in the range
- \( n \) is sample size

The chi-square test is defined for the hypothesis to test the GoF is:

\( H_0: \) The data follow a specified distribution.

\( H_a: \) The data do not follow the specified distribution.

The hypothesis that the data are from a population with the specified
distribution is rejected if

\[ \chi^2 \geq \chi^2_{\alpha,k} \]

where \( \chi^2_{\alpha,k} \) is the chi-square critical value with \( k \) degrees of freedom and significance
level \( \alpha \).
5. FITTING OF AN EXPONENTIAL DISTRIBUTION

5.1 Probability density function

The probability density function (pdf) of an exponential distribution is

\[
f(x; \lambda) = \begin{cases} 
\lambda e^{-\lambda x}, & x \geq 0, \\
0, & x < 0.
\end{cases}
\]  

where \( \lambda \) is the parameter.

The parameters are usually unknown and are to be estimated from the data. It is also necessary to check whether the distribution of observed data fits into one of the known theoretical distributions or not, so that the distribution can be better understood, if necessary, for better control and managerial efficiency.

5.2 Cumulative Distribution Function (CDF)

The cumulative distribution function (cdf) of exponential distribution is given by;

\[
F(x; \lambda) = \begin{cases} 
1 - e^{-\lambda x}, & x \geq 0, \\
0, & x < 0.
\end{cases}
\]  

The mean and variance of the exponential distribution are given by:

\[
Mean = \frac{1}{\lambda}; \quad Variance = \frac{1}{\lambda^2};
\]  

The collected data (Table No.1) has been used to develop a model to fit the data of power units and tested good fit to the repair hours for all seven units i.e. Unit-1, Unit-2, Unit-3, Unit-4, Unit-5, Unit-6 and Unit-7 for the period of seven years.

<table>
<thead>
<tr>
<th>Table No. 1: Repair Hours of Unit 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
</tr>
<tr>
<td>0.09</td>
</tr>
<tr>
<td>0.14</td>
</tr>
<tr>
<td>0.17</td>
</tr>
<tr>
<td>0.22</td>
</tr>
<tr>
<td>0.32</td>
</tr>
<tr>
<td>0.57</td>
</tr>
<tr>
<td>0.97</td>
</tr>
<tr>
<td>2.02</td>
</tr>
<tr>
<td>2.91</td>
</tr>
<tr>
<td>7.34</td>
</tr>
<tr>
<td>14.04</td>
</tr>
</tbody>
</table>
C.P.S. Hungund and Shrikant Patil

The procedure and calculation has been given only for one unit i.e. Unit-1. The same procedure is followed to all other units. Here from data collected, assessed the exponentiality of the data tested via the Chi-Square GoF test.

Firstly descriptive statistics are needed to be calculated (Table No. 2). Using these, we can define the subintervals.

**Table No. 2: Descriptive Statistics.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Lambda</th>
<th>Median</th>
<th>StdDev</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>84</td>
<td>3.44</td>
<td>0.29</td>
<td>0.57</td>
<td>5.89</td>
<td>0.02</td>
<td>24.35</td>
<td>0.17</td>
<td>2.88</td>
</tr>
</tbody>
</table>

The hypothesis of the problem is:

- $H_0$: The data follows Exponential Distribution.
- $H_a$: The data do not follow the Exponential Distribution.

For the above, data ($\lambda=3.44$) is generated on exponential distribution assessment. Now we will assess the exponentiality of the data via the Chi-Square GoF test.

For endpoints we now select 0.10, 0.15, 0.20, 0.25, 0.50, 0.60, 1.5, 2.5, 3.0, 6.5, 9.5, 20 and 25 which in turn define subintervals. We obtain cumulative and individual cell probability values.

For Ex,

$$P_{3.44}(0.10) = 1.0 - \exp\left[-\frac{0.10}{3.44}\right] = 0.03$$

The entire GOF process for this case is summarized in the following table:

**Table No. 3: Intermediate Values for the exponential distribution GoF test.**

<table>
<thead>
<tr>
<th>Intermediate Value (X)</th>
<th>f</th>
<th>Cum Prob</th>
<th>Cell Prob</th>
<th>Expected</th>
<th>Observed</th>
<th>Pool</th>
<th>Pooled Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chi-Square</td>
</tr>
<tr>
<td>0.1</td>
<td>8</td>
<td>0.03</td>
<td>0.03</td>
<td>2.41</td>
<td>8</td>
<td>3</td>
<td>5.89</td>
</tr>
<tr>
<td>0.15</td>
<td>7</td>
<td>0.04</td>
<td>0.01</td>
<td>1.18</td>
<td>7</td>
<td>8</td>
<td>5.48</td>
</tr>
<tr>
<td>0.2</td>
<td>9</td>
<td>0.06</td>
<td>0.01</td>
<td>1.16</td>
<td>9</td>
<td>1</td>
<td>18.32</td>
</tr>
<tr>
<td>0.25</td>
<td>8</td>
<td>0.07</td>
<td>0.01</td>
<td>1.14</td>
<td>8</td>
<td>7</td>
<td>13.7</td>
</tr>
<tr>
<td>0.5</td>
<td>8</td>
<td>0.14</td>
<td>0.07</td>
<td>5.48</td>
<td>8</td>
<td>4</td>
<td>5.49</td>
</tr>
<tr>
<td>0.6</td>
<td>5</td>
<td>0.16</td>
<td>0.02</td>
<td>2.08</td>
<td>5</td>
<td>4</td>
<td>22.42</td>
</tr>
<tr>
<td>1.5</td>
<td>8</td>
<td>0.35</td>
<td>0.19</td>
<td>16.24</td>
<td>8</td>
<td>7</td>
<td>7.39</td>
</tr>
<tr>
<td>2.5</td>
<td>7</td>
<td>0.52</td>
<td>0.16</td>
<td>13.7</td>
<td>7</td>
<td>9</td>
<td>5.06</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.58</td>
<td>0.07</td>
<td>5.49</td>
<td>4</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>6.5</td>
<td>4</td>
<td>0.85</td>
<td>0.27</td>
<td>22.42</td>
<td>4</td>
<td>8</td>
<td>84</td>
</tr>
<tr>
<td>9.5</td>
<td>7</td>
<td>0.94</td>
<td>0.09</td>
<td>7.39</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>1</td>
<td>0.06</td>
<td>5.06</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The result of the Chi-Square Goodness of Fit test statistic for this data is:

\[
\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 140.36
\]

The critical value at 5% level of significance and 7 df is 14.07. Since the tabulated value is greater than the critical value, we reject H0 and conclude that the exponential distribution is not a good fit to the data.

The procedure explained above to obtain GoF is for the Unit-1. Similar procedure can be applied to other units also. Results obtained for other units are along with Unit-1 are given in Table No. 4 along with their chi value to test the GoF.

**Table No. 4: Summary of the exponential distribution of Repair Hours.**

<table>
<thead>
<tr>
<th>Units</th>
<th>n</th>
<th>k</th>
<th>Mean</th>
<th>Lamda</th>
<th>Chi-Square</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>84</td>
<td>13</td>
<td>3.44</td>
<td>0.29</td>
<td>140.36</td>
<td>7</td>
</tr>
<tr>
<td>Unit 2</td>
<td>100</td>
<td>12</td>
<td>2.48</td>
<td>0.4</td>
<td>116.5</td>
<td>6</td>
</tr>
<tr>
<td>Unit 3</td>
<td>60</td>
<td>6</td>
<td>3.03</td>
<td>0.33</td>
<td>120.34</td>
<td>3</td>
</tr>
<tr>
<td>Unit 4</td>
<td>60</td>
<td>6</td>
<td>1.98</td>
<td>0.505</td>
<td>85.56</td>
<td>2</td>
</tr>
<tr>
<td>Unit 5</td>
<td>56</td>
<td>8</td>
<td>3.14</td>
<td>0.318</td>
<td>46.07</td>
<td>4</td>
</tr>
<tr>
<td>Unit 6</td>
<td>51</td>
<td>6</td>
<td>3.1</td>
<td>0.323</td>
<td>37.43</td>
<td>3</td>
</tr>
<tr>
<td>Unit 7</td>
<td>82</td>
<td>8</td>
<td>3.05</td>
<td>0.328</td>
<td>143.85</td>
<td>4</td>
</tr>
</tbody>
</table>

6. ANALYSIS

Table No. 2 and 3 gives the result obtained for estimating the statistical parameter for Repair Hours of Unit-1. Repair Time data are fitted with exponential distribution. The applicability of exponential distribution was examined and inferred that none of them give a good fit for repair time data of Unit-1 at both 5% and 1% level of significance and also for all other units.

Here though the analysis has not given good fit to the data, but investigation enables us to identify if the parameter λ is a representative of the data. Since there exist large scatteredness of the observation about the parameter value, the data values are more deviated, thereby implying that the distribution is not a good fit to the data. In order to minimize variation about the parameter, data needs to be classified on certain constraints like equipment wise, duration wise. Accordingly in the next paper, focus will be on classifying data into two categories; so that the procedure used to develop the model may give a good fit to the data. And also it can be shown that exponential distribution is a good fit to the data when failure rate λ is constant.

7. CONCLUSION
From the Table No. 4, we observe that none of the units have not found good with exponential distribution. More importantly we can conclude that fitting of repair hours for the data related to power sector necessarily requires data to be divided into categories based on some constraints.

REMARKS

The above procedure can also be extended to show that the Weibull and Lognormal distributions also give a good fit for the observed failure and working times of power units.

ACKNOWLEDGEMENT

Authors sincere thanks to Dr. K.L.A.P. Sharma, Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapur; for his useful suggestion and valuable comments in the presentation of the paper.

REFERENCES

Failure Analysis of Thermal Power Plant Using Normal and Lognormal Distributions

C.P.S. Hungund*, Shrikant Patil**

{*Professor, **Research Scholar} Department of Statistics, Gulbarga University, Gulbarga, Karnataka, INDIA.

Corresponding Addresses:
cpshgu@gmail.com,** shrikant07.inc@gmail.com

Research Article

Abstract: The authors have analyzed the failure patterns of power units of Thermal Power Plant located at Shakti Nagar, Raichur, Karnataka. In the study, it was found that Weibull distribution is good fit to the working hour data compared with exponential distribution [3]-[5]. But the study would be incomplete unless we test it through Normal and Lognormal distributions. The reasons may be Normal and Lognormal are widely used distributions in reliability studies and together they are addressed because they are intimately related. It is always our interest in assessing whether a data set comes from such distributions. Consequently, when a random variable is distributed Lognormal, the Logarithm (base e) of the random variable is distributed Normal. This property carries our data sets too. In this paper, we have made an attempt to fit Normal and Lognormal distributions to power units of failure data related to thermal power plant. For the collected data, Normal and Lognormal Models are fitted and tested for its Goodness of Fit (GoF). Based on the fitted model we have estimated reliabilities of thermal power units. Conclusions were drawn based on the result obtained.

Keywords: Thermal Power Plant, Working Hours, Normal and Lognormal Distribution, Reliability.

1. Introduction
Power sector being one of the prominent industrial sectors, play a vital role in the economic growth of the country. Key players of power sector can be listed as Thermal Power Plant (TPP), Hydropower or Hydroelectricity, Solar Power, Biogas Energy, Wind Power and so on. Among these, thermal power generating units’ accounts 68.14 % compared with other sources of power units in India. The overall plant load factor of the thermal power generating units is about 70.9 %. A little briefing can be done from above statistics; (i) electric generation in India right now is in critical state; (ii) since thermal power plants (TPP) accounts more than any other source of power generation, failure free operation of the TPP is essential to achieve economic growth of the country. Hence to meet the demand supply gap, apart from augmenting the capacity, there is an immense need to improve the performance of the existing power generating units. Therefore it has become essential for us to understand the failure patterns or downtime of the production units, to sustain failure free operations. In this study, we have made an analysis of failure patterns of TPP through probabilistic approach using Normal and Lognormal distribution and assessing the reliabilities of power units using reliability theory. For the present study, we have chosen Raichur Thermal Power Station (RTPS), Shaktinagar located about 20 Km from Raichur, Karnataka State to study thermal units. A unit represents the system consisting of Steam generator, Boiler furnace and steam drum, Superheater, Reheater, Steam turbine, Auxiliary systems, Fuel preparation system etc., involving in electricity productions. There are such seven power generating units (Unit-1 to Unit-7) at RTPS. The data on working hours in seven power generating units of thermal power plant are collected during the period 2004-2011.

2. Survey of Literature
D. D. Adhikary, G. K. Bose, S. Chattopadhyay, D. Bose and S. Mitra [2] dealt a case of “RAM investigation of coal-fired thermal power plants: A case study”. The authors have investigated the reliability, availability and maintainability (RAM) characteristics of a 210 MW coal-fired thermal power plant (Unit-2) from a thermal power station in eastern region of India. Analyses of components/equipments have been tested through Weibull and Lognormal distribution, later GoF test have been performed through Kolmogorov–Smirnov Test. Critical mechanical subsystems with respect to failure frequency, reliability and maintainability are identified for taking necessary measures for enhancing availability of the power plant and the results are compared with the same Power Station. The author concludes that RAM analysis is very much effective in finding critical subsystems and deciding their preventive maintenance program for improving availability of the power plant as well as the power supply. In the paper “Reliability Analysis of Thermal Power Generating Units based on Working Hours” [4] by Hungund CPS and Shrikant Patil, the authors have considered seven years data on working hours for testing the suitability of the exponential and Weibull distribution. The applicability of the distributions for working hours has been tested through chi-square test.
of GoF. The test reveals that weibull distribution is the most reliable distribution for working hours to be used for fitting the data. Later reliability analysis of weibull and exponential performed to identify the best and poor performing units. Finally the authors made a remark that by taking necessary measures the reliabilities of poor units can be enhanced. Romeu J.L. [11] discussed some empirical and practical methods for checking and verifying the statistical assumptions of Normal and Lognormal distributions in the paper “Empirical Assessment of Normal and Lognormal Distribution Assumptions”. In the study, two distribution assumptions were verified: (i) the data are independent and (ii) they are identically distributed as a Normal. Later these assumptions were verified through the important properties normal distribution. These properties carried on to the even Lognormal distribution too, since when a data set comes from a Lognormal population, then the logarithm of these data are distributed as a normal. With the use of different examples and graphs, the procedure was shown to verify the assumption of normality data set.

3. Objectives
The objectives of the present study are:
• To fit the normal and lognormal distribution for working hours.
• To test the assumption of the normal and lognormal through the chi-square test of GoF.
• To perform the reliability analysis of seven power units of a thermal power plant.

4. Methodology
4.1 Verification of Normal Assumption
Numerous ways of verifying distribution assumption are: distribution properties, graph, and GoF test. The randomization of population units need to be verified before we put them for test to show data independence and identical. The time of operation, operator, units, weather conditions etc are considered randomly for the study to represent same characteristics in which unit are operating normally.

The properties of normal distributions are used for easy assessment of normality of the data. The properties are:
1. Mean, Median and Mode coincide; hence, sample values should also be close.
2. Graph should suggest that the distribution is symmetric about the mean.
3. Should satisfy 68-95-99.7 rule or empirical rule.
4. Plots of the Normal Probability and Normal scores should be close to linear.

Firstly, we put the collected data of unit-1 in the following table. Note that the logarithms of actual data are taken for consideration.

Table 1: Logarithmic of working hours of Unit-1(Sorted)

<table>
<thead>
<tr>
<th>Value</th>
<th>-1.9741</th>
<th>-1.9301</th>
<th>-1.2274</th>
<th>-0.5704</th>
<th>0.0000</th>
<th>0.0000</th>
<th>0.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0079</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
<tr>
<td>0.6931</td>
<td>0.6931</td>
<td>0.6931</td>
<td>1.0986</td>
<td>1.0986</td>
<td>1.1244</td>
<td>1.3863</td>
<td>1.3863</td>
</tr>
<tr>
<td>1.3863</td>
<td>1.3863</td>
<td>1.3863</td>
<td>1.4035</td>
<td>1.6094</td>
<td>1.7918</td>
<td>2.0794</td>
<td>2.0794</td>
</tr>
<tr>
<td>2.3026</td>
<td>2.3026</td>
<td>2.3026</td>
<td>2.3026</td>
<td>2.3026</td>
<td>2.3071</td>
<td>2.4849</td>
<td>2.4849</td>
</tr>
<tr>
<td>2.4849</td>
<td>2.5649</td>
<td>2.5649</td>
<td>2.6956</td>
<td>2.7726</td>
<td>2.8332</td>
<td>2.8332</td>
<td>2.8332</td>
</tr>
<tr>
<td>4.6913</td>
<td>4.7829</td>
<td>5.0903</td>
<td>5.2979</td>
<td>5.2979</td>
<td>5.2979</td>
<td>5.2979</td>
<td>5.2979</td>
</tr>
</tbody>
</table>

To assess the data, we obtain their descriptive statistics, and then analyze and plot the raw data to check if the Normality assumption holds.

Table 2: Descriptive Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>74</td>
<td>2.2669</td>
<td>2.5649</td>
<td>1.6731</td>
<td>1.1051</td>
<td>3.4257</td>
</tr>
</tbody>
</table>

From table no.2, the sample mean (2.2669) and Median (2.5649) are close. This supports the normality of the distribution by property no. 1. The distribution looks symmetric about mean (2.2669) since approximately 52% of the centered data between Q1 (1.1051) and Q3 (3.4257). Also note that the highest frequency lies in between data 1.00 and 3.00. All of these by property no. 2, suggests the validity of the normal distribution. The interval defined by one standard deviation about the mean: (µ-σ, µ+σ) = (0.5938, 3.9400) includes 50 values representing 67.57% of the total data set close to the expected 68.25%. The interval (µ-2σ, µ+2σ) = (-1.0793, 5.6131) includes 72 values representing 97.30% of the data set (close to the expected of 95%). There are zero values beyond µ±3σ, supporting the statement that about 1% of the values would be outside the interval (µ-3σ, µ+3σ). These results support the empirical rule. In the probability plot (Figure-1), the normal probability is plotted against I/(n+1) where I is the data sequence order, i.e. I=1,2,…..74. Each P_i is obtained by calculating the
Normal probability of the corresponding failure data, $X_i$ using the sample mean and the standard deviation as shown in the following equation.

$$P_{\mu,\sigma}(x) = \text{Normal} \left( \frac{x - \mu}{\sigma} \right)$$

For instance, at $I=1$, the data point is -1.9741:

$$P_{-1.9741, 1.6731}(X) = \text{Normal}(-2.5348) = 0.0056$$

Similarly, $P_i$ is computed for each of the data sequence in the set. The data point is then plotted against the corresponding $I/(n+1)$ until done with all sample elements. The corresponding figure is shown below:

![Figure 1: Normal Probability Plot](image)

When the population is normal, the probability plot shown in above figure-2 follows an upward linear trend. The regression index of fit $R^2=97.62\%$ is very high close to 100% suggesting the linear trend. This serves the property no. 4.

From the above discussions, the assumptions of normality holds true; in the following section we fit the data using normal and lognormal distributions and test it through by implementing chi-square test of GoF. Later reliability analysis is carried out to know the best performing units at RTPS.

## 5. Goodness of Fit Test

The failures of thermal units are due to various reasons. In order to identify the best suitable analysis, we use the procedure suggested by Hungund CPS et. al., (2003). In the paper the authors developed a model to fit the data of power units to the working hours for all seven units.

### 5.1 Fitting of Lognormal Distribution

The lognormal distribution is one of the most widely used distributions of time to failure. The distribution is commonly used to model the lives of the units whose failure modes are fatigue in nature. Due to this fact, the lognormal distribution has widespread application. Most of the time, the lognormal distribution is used along with the Weibull distribution when attempting to model failure of units. The lognormal distribution has certain similarities to the normal distribution. A random variable is lognormally distributed if the logarithm of the random variable is normally distributed. Because of this fact, there are many mathematical similarities between the two distributions.

#### 5.1.1 Probability Density Function

The lognormal is denoted by that name since, if $X$ is the random variable representing the lognormal time to failure, the random variable, $Y = \ln X$, is normally distributed with parameters mean $\mu$ and standard deviation $\sigma$ where $\mu>0$ and $\sigma>0$. If the probability density function of $X$ is:

$$f_X(x) = \frac{1}{\alpha \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (\ln(x) - \mu)^2}$$

for $x>0$.

#### 5.1.2 Cumulative Distribution Function

The cumulative distribution function (cdf) of lognormal distribution is given by:

$$F_X(x) = F \left( \frac{\ln(x) - \mu}{\sigma} \right)$$

Where $F(z)$ is the cumulative probability distribution function of $N(0,1)$.

#### 5.1.3. Testing of Hypothesis

The hypothesis of the problem is given by:

$$H_0: \text{The data follows lognormal distribution.}$$

$$H_a: \text{The data do not follow the lognormal distribution.}$$

We obtain the point estimations of the assumed lognormal distribution parameters mean and standard deviation. The point estimations allow us to define the composite distribution hypothesis $\mu=2.2669$ & $\sigma=1.6731$. Since parameters mean and variance were estimated from the data the resulting chi-square statistic degrees of freedom are: $df=k-2-1$.

For endpoints (Table No. 1) we now select 0.0000, 0.4055, 1.3863, 2.3026, 2.8332, 3.1355, 3.4340, 3.8712, 4.4427 and 5.2983 which in turn define subintervals. We obtain cumulative and individual cell probability values. For

$$P_{2.2669, 1.6731}(X) = \text{Normal}(0, 2.2669) = \text{Normal}(1.3549) = 0.0877$$

### Table 3: Intermediate Values for the lognormal distribution GoF test

<table>
<thead>
<tr>
<th>$X$</th>
<th>StdEnd</th>
<th>CumProb</th>
<th>CellProb</th>
<th>Ei</th>
<th>Oi</th>
<th>Ei</th>
<th>Oi</th>
<th>Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1.3549</td>
<td>0.0877</td>
<td>0.0874</td>
<td>6.4676</td>
<td>7</td>
<td>9.8118</td>
<td>14</td>
<td>1.7877</td>
</tr>
<tr>
<td>0.4055</td>
<td>-1.1126</td>
<td>0.1329</td>
<td>0.0452</td>
<td>3.3465</td>
<td>7</td>
<td>12.3175</td>
<td>10</td>
<td>0.436</td>
</tr>
<tr>
<td>1.3863</td>
<td>-0.5263</td>
<td>0.2993</td>
<td>0.1664</td>
<td>12.3119</td>
<td>10</td>
<td>15.4978</td>
<td>9</td>
<td>2.7244</td>
</tr>
<tr>
<td>2.3026</td>
<td>0.0213</td>
<td>0.5085</td>
<td>0.2092</td>
<td>15.4788</td>
<td>9</td>
<td>9.187</td>
<td>9</td>
<td>0.0038</td>
</tr>
</tbody>
</table>
The density functions of these distributions are frequently used, the Weibull, the Gamma and Normal distributions are the accepted choices for time to failure (TTF) distributions with monotone hazard or failure rates. The density of the normal distribution is a good fit to the data.

The normal distribution is chosen as models for the frequency of occurrence of TTF values. This choice is made because either their theoretical properties are consistent with the conditions of use and the physics of failure of the device, or because the density adequately describes the failure history of the device. Often, the normal and lognormal distributions are addressed together because they are closely related.

5.2.1. Probability Density Function
A random variable $X$ is said to have a normal (or Gaussian) distribution with parameters $\mu$ and $\sigma$, where $-\infty < \mu < \infty$ and $\sigma > 0$, with probability density function:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < \infty, \quad -\infty < \mu < \infty,$$

5.2.2 Normal cumulative distribution function

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(t-\mu)^2}{2\sigma^2}\right) dt = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

5.2.3. Testing of Hypothesis
The hypothesis of the problem is given by:

$H_0$: The data follow normal distribution.

$H_a$: The data do not follow the normal distribution.

Here we use similar procedure which was used earlier in case of lognormal distribution. Since normal distribution does not satisfy many of the properties stated earlier, therefore we will not elaborate much about normal distribution. However, results obtained after testing chi-square test of GoF.

**Table 5:** Summary of the Normal distribution for working Hours.

<table>
<thead>
<tr>
<th>Units</th>
<th>n</th>
<th>k</th>
<th>Mean</th>
<th>StdDev</th>
<th>Chi-Square</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>74</td>
<td>10</td>
<td>2.2669</td>
<td>1.6731</td>
<td>8.361</td>
<td>5</td>
</tr>
<tr>
<td>Unit 2</td>
<td>94</td>
<td>11</td>
<td>2.0834</td>
<td>1.3820</td>
<td>4.582</td>
<td>7</td>
</tr>
<tr>
<td>Unit 3</td>
<td>56</td>
<td>7</td>
<td>2.0820</td>
<td>1.4380</td>
<td>2.125</td>
<td>3</td>
</tr>
<tr>
<td>Unit 4</td>
<td>59</td>
<td>9</td>
<td>2.2697</td>
<td>1.4970</td>
<td>2.125</td>
<td>3</td>
</tr>
<tr>
<td>Unit 5</td>
<td>52</td>
<td>8</td>
<td>2.3301</td>
<td>1.6420</td>
<td>1.583</td>
<td>3</td>
</tr>
<tr>
<td>Unit 6</td>
<td>46</td>
<td>6</td>
<td>2.0932</td>
<td>1.6649</td>
<td>1.583</td>
<td>3</td>
</tr>
<tr>
<td>Unit 7</td>
<td>79</td>
<td>8</td>
<td>2.0690</td>
<td>1.6298</td>
<td>3.538</td>
<td>5</td>
</tr>
</tbody>
</table>

5.3 Analysis

Working time data are fitted with Lognormal and Normal distributions. Table No. 4 & 5 gives the result obtained after performing chi-square test of GoF for working hours of seven units. From the table no. 4, it can be seen that the chi-square values are accepted both at 5% and 1% level of significance. Therefore it is inferred that lognormal distribution is a good fit to the data on working hours for all the units. It is interesting to note that, none of the units are accepted in case of normal distribution (Table No. 5).

Since normal distribution did not satisfy the distribution assumptions, but still we checked it through chi-square test of GoF, However $H_0$ is rejected for all the units. Yet another numerically convoluted proof that, the normal distribution is not a good fit to the data related to thermal power units.

6. Reliability Analysis

In general terms, reliability is “the ability of an entity to perform required function under given conditions for a given period of time”. In technical terms, reliability is measured by the probability that a system or a component will work without failure during a specified time interval under given operating conditions. The term reliability can be applied to almost any object, which is the reason that the term system, equipment and component are used in the definition. Reliability is defined positively, in terms of a system performing its intended function, and no distinction is made between failures. Nevertheless, for system reliability analysis, there must be a great deal of
concern not only with the probability of failure but also with the potential consequences of failures that present severe safety and economic loss or inconvenience. The reliability analysis is performed for each of the thermal units in the power plant. The reliability analysis is based on the time to failure data analysis. To better understand the behavior of lifetime distributions, reliabilities of lognormal analysis is performed in the following sections.

6.1 Reliability of Lognormal Distribution
The lognormal distribution is one of the most widely used in reliability problems. As with the normal distribution, there is no closed-form solution for the lognormal reliability function. Solutions can be obtained via the use of standard normal tables. Since the application automatically solves for the reliability. The reliability function is given below:

$$R(t) = 1 - F \left( \frac{\ln(t) - \mu}{\sigma} \right)$$

The reliability distribution for lognormal distribution curve for all the units are presented in table no. 5 as well as figure-3. The points presented in the graph represent the reliability estimate for each of the time to failure data, arranged in increasing order with time t. Those points are used to verify the adherence of the reliability distribution to the failure data.

| Table 6: Reliability of the seven units of Thermal Power Plant at different time unit 't'. |
|----------------------------------|---------|---------|---------|---------|---------|---------|---------|
|                                | Unit1   | Unit2   | Unit3   | Unit4   | Unit5   | Unit6   | Unit7   |
| Mean                            | 2.2669  | 1.0834  | 2.3908  | 2.6129  | 2.9757  | 2.9932  | 2.3690  |
| SD                              | 1.6731  | 2.2337  | 2.0820  | 1.8578  | 1.3904  | 1.6649  | 1.6298  |
| 5                               | 0.6528  | 0.4065  | 0.6463  | 0.7054  | 0.8371  | 0.7971  | 0.6794  |
| 10                              | 0.4915  | 0.2918  | 0.5169  | 0.5663  | 0.6858  | 0.6609  | 0.5163  |
| 15                              | 0.3960  | 0.2325  | 0.4394  | 0.4796  | 0.5763  | 0.5680  | 0.4176  |
| 20                              | 0.3316  | 0.1949  | 0.3857  | 0.4184  | 0.4943  | 0.4994  | 0.3503  |
| 25                              | 0.2847  | 0.1684  | 0.3454  | 0.3721  | 0.4306  | 0.4461  | 0.3010  |
| 50                              | 0.1627  | 0.1017  | 0.2325  | 0.2422  | 0.2503  | 0.2905  | 0.1719  |
| 100                             | 0.0811  | 0.0566  | 0.1438  | 0.1418  | 0.1206  | 0.1665  | 0.0850  |
| 150                             | 0.0505  | 0.0387  | 0.1041  | 0.0984  | 0.0717  | 0.1128  | 0.0525  |
| 200                             | 0.0350  | 0.0290  | 0.0813  | 0.0742  | 0.0474  | 0.0831  | 0.0361  |
| 300                             | 0.0200  | 0.0189  | 0.0558  | 0.0481  | 0.0249  | 0.0518  | 0.0204  |

6.2 Analysis
From the above table and graph is clearly seen that Unit-5 is slightly shows higher reliability than unit-6 initially but later (after 20th hour) unit-6 starts showing good reliability than Unit-5. This concludes that unit-6 is best performing unit among the other six units. Note that reliability of power units after 100 hours. Lognormal reliability is below 0.20 mark and reliability is almost closer to zero mark at 150th hours except unit-5 and unit-6. Also unit-2 is showing lesser reliability in the distribution when compared with other units which indicates necessary measure has to be taken for improvement.

7. Conclusion
The reliability analysis of thermal power generating units based on working hours have been tested using lognormal distribution. Later GoF have been performed through chi-square test for both normal and Lognormal distribution. The GoF test reveals that lognormal distribution is the most reliable distribution for working hours to be used for fitting the data. Reliabilities are identified for taking
necessary measures enhancing availability of the power plant.

Acknowledgement
Authors sincere thanks to Dr. K.L.A.P. Sarma (Retd), Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapur; for his useful suggestion and valuable comments in the presentation of this paper.

References
TO APPEAR IN THE RESEARCH JOURNALS
ESTIMATION OF RELIABILITIES OF THERMAL POWER PLANT – A CASE STUDY

C. P. S. Hungund* and Shrikant Patil
Department of Statistics, Gulbarga University, Gulbarga, India.
e-mail: cpshgu@gmail.com

Abstract: To sustain smoother flow of electricity generation of thermal power plant, it is very important to understand the pattern of successive failures (working hours) and production downtime (repair hours). Since exponential distribution is a commonly assumed model for working hours and repair hours, the applicability of exponential distribution is tested through Chi-Square test of Goodness of Fit (GoF). Also an attempt is made to find out the best suitable analysis for fitting production downtimes.

Key words: Working hours and repair hours, Exponential distribution, Chi-square test of GoF.

1. Introduction
An operation of power station with the highest possible availability has made it extremely important in recent years; the reliability of a thermal power station obviously has a vital influence in this respect. In addition, the liberalization of the electric power systems pushes the operators of power stations in the direction of achieving the highest availability at the lowest costs. This paper is specifically concerned with the reliability and failure assessment of thermal power stations. Repair Hour (Production Downtime) data obtained from the operating history (i.e. from the year 2004 to 2011) of Raichur Thermal Power Station (RTPS) located at Shaktinagar, Raichur have been used for the investigation. The reliability of the power stations has been assessed by means of Probability Distributions. The probability distributions take into account the repair hours, the average repair hour and the number of thermal units. The system reliability is the most important concern and generation of the power mainly depends on the effective functioning of the system. In order to estimate the reliability of a system one must concentrate on the failure patterns or downtime of the production units.

In the paper, “Analysis of Production Downtimes in Thermal Power Plant using Exponential Failure Model A Case Study” Hungund C.P.S. & Shrikant P. (2013), we have developed a model based on failure patterns of seven thermal power units located at RTPS, Raichur. Exponential model is used to fit the data through Chi-square test of GoF. It has been observed that there exists large scatteredness of observation about the parametric value; the data values are more deviated thereby implies that, the distribution is not good fit to the data. In order to minimize the variation about the parameter and assumed distribution give a good fit to the data, data need to be classified on duration wise so that the procedure used to develop the model may give good fit to the data.

The objective of present paper is to identify the best suitable analysis to develop a model to fit the data of power units of RTPS, Raichur. The focus will be on classifying data based on major repair hours and minor repair hours so that the procedure is used to develop the model may give good fit to the data. The Chi-square test is applied to know the underlying the distribution supports data or not.

In the following sections we have discussed survey of literature, objectives of the study, methodology, fitting of Exponential Distribution for repair hours of seven power units of thermal power plant at Raichur, analysis and conclusion.

2. Survey of Literature
Braun et al. (2003) presented a paper on Reliability and Economic Analysis of Different Power Station Layouts, at IEEE Bologna Power Tech Conference, Bologna, Italy. This paper is specifically concerned with the reliability and security assessment of thermal power stations. Equipment Failure and repair data obtained from the two Italian power stations have been used for the investigation. In order to prove that, assumption of the
data is exponentially distributed, the authors have shown that the parameter $\lambda$ is constant. The initial step was to analyze the data to compute a piecewise-continuous failure density function. A study of this function is then followed by the choice of a continuous model which fits the data satisfactorily. A hazard rate is a measure of the instantaneous speed of failure. A constant hazard rate implies an exponential density function and an exponential reliability function. The authors have proven a constant hazard rate and thus concluded that exponential distribution is a good fit to the data relating to thermal power unit.

Romeu (2004) discussed Goodness of Fit test using various distributions in “The Chi-Square: a Large Sample Goodness of Fit Test”. In the study, the author has developed a model to fit the data using Normal Distribution, Exponential Distribution and Weibull Distribution. Then Chi-square test was applied to know the underlying distribution supports the data or not. The author has taken several examples for Normal, Lognormal, Exponential and Weibull distributions and showed how the Goodness of Fit can be dealt via Chi-square test. The paper discusses importance of GoF assessments of statistical distributions. Also showed that, for the small sample GoF cannot be done through Chi-square test. For the number of observations per cell is too small for the GoF test statistic to converge to its Chi-square underlying distribution. In such cases, one can use Cumulative Distribution Function (CDF) based distance Goodness of Fit tests, such as the Anderson-Darling and Kolmogorov-Smirnov test.

Reddy (1992) dealt a case of “Modeling the Causes of Production Downtimes: An Empirical Study of a Thermal Power Generating Unit the Causes of Production Downtimes”. In the study, the author has studied the causes of production downtimes in thermal power generation. An attempt is made to find which distribution is best fitted for working time between successive failures (working hours) and downtime (repair hours). Exponential distribution is used as a model of failure time and repair time. Chi-square test is used to test the GoF.

3. Objectives

The objectives of this study are:

- To fit the exponential distribution for repair hours of seven power units of a thermal power plant.
- To test the assumption of exponentiality through the chi-square test of GoF.
- To identify best suitable analysis to analyze repair hours of thermal power plant.

4. Methodology

The data on repair hours in seven power generating (namely Unit-1, Unit-2, Unit-3, Unit-4, Unit-5, Unit-6 and Unit-7) of thermal power plant are collected for a period of seven years. The failures of thermal units are due to various reasons like failure of Electrical Turbine, Boiler, Furnace, Graft Failure etc. An attempt is made to find out suitable distribution for downtimes. Exponential distribution is commonly assumed model of time to repair hours.

Establishing the underlying distribution of a data set or random variable is crucial for the correct implementation of some statistical procedures. For example, deriving the test and Class Interval (CI) for the population Mean Time Between Failure (MTBF) requires knowledge about the distribution of the lives of the device. If the lives are exponential, things will be done one way; if they are Weibull, they will be done differently. Therefore, first need to establish the failure distribution from the data, before one can correctly implement the test procedures.

GoF tests are essentially based on either of two distribution basics: the Cumulative Distribution Function (CDF), and the Probability Density Function (PDF). Procedures based on the CDF are called “distance tests” while those based on the PDF are called “area tests”.

To assess data, we implement a well-defined scheme. First, assume that data follow a pre-specified distribution. Then, we either estimate the distribution parameters (e.g., mean and variance) from the data or obtained from prior experience. Such process yields the “composite” distribution hypothesis called the null hypothesis. The negation of the assumed distribution is called the alternative hypothesis. We then test the assumed distribution using the data set. Finally, $H_0$ is rejected whenever any one (or more) of the several elements in hypothesis $H_0$ is not supported by the data.

The Chi-square test is conceptually based on the probability density function (PDF) of the assumed distribution. If this distribution is correct, its PDF should closely encompass the data range (of X). We thus select convenient values in this data range that divide it into several subintervals. Then, we compute the number of data points in each subinterval. These are called “observed” values. Then, we compute the number that should have fallen in these same subintervals, according to the PDF of the assumed distribution. These are called the “expected” values and the Chi-square test requires at least five of them in every subinterval. Finally, we compare these two results. If they agree (probabilistically)
then the data supports the assumed distribution. If they do not, the assumption is rejected. The formula (statistic) that uses the differences between “expected” and “observed” values to test the Goodness of Fit follows a Chi-square distribution. Hence, the name Chi-square test.

The following are the step-by-step summary of Chi-square test of GoF:

1. Establish the null hypothesis.
2. Estimate the exponential parameter from the data.
3. Establish the k number of subintervals.
4. Obtain probability for the K number of subintervals.
5. Test statistic distribution: chi-square; DF.
6. Establish significance level \( \alpha = 0.05 \).
7. Obtain chi-square critical value.
8. Compute test statistic value.
9. As critical value > test statistic, assume exponentially.

The distribution of test statistic is as follows:

\[
\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]

Where, \( E_i \) is expected number of data points (\( E_i \geq 5 \))
\( O_i \) is observed number of data points
\( k \) is total number of cell or subintervals in the range
\( n \) is sample size

The chi-square test is defined for the hypothesis to test the GoF is:

\( H_0 \): The data follow a specified distribution.
\( H_a \): The data do not follow the specified distribution.

The hypothesis that the data are from a population with the specified distribution is rejected if

Where, \( \chi^2_{\alpha,k} \) is the chi-square critical value with \( k \) degrees of freedom and significance level.

5. Fitting of an Exponential Distribution

5.1 Probability density function

The probability density function (pdf) of an exponential distribution is

\[
f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}
\]

Where, \( \lambda \) is the parameter.

The parameters are usually unknown and are to be estimated from the data. It is also necessary to check whether the distribution of observed data fits into one of the known theoretical distributions or not, so that the distribution can be better understood if necessary, for better control and managerial efficiency.

5.2 Cumulative Distribution Function (CDF)

The cumulative distribution function (cdf) of exponential distribution is given by

\[
F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}
\]

The mean and variance of the exponential distribution are given by

\[
\text{Mean} = \frac{1}{\lambda}; \quad \text{Variance} = \frac{1}{\lambda^2};
\]

In order to identify the best suitable analysis, the collected data has been split into two categories viz., Category-I and Category-II. Category-I represents minor repairs hours and category-II represents major repair hours; so that the procedure suggested by Romeu (2004), is used to develop a model to fit the data of power units and gave a good fit to the repair hours for all seven units i.e. Unit-1, Unit-2, Unit-3, Unit-4, Unit-5, Unit-6 and Unit-7 for the period of seven years.

The procedure and calculation has been given only for one unit \textit{i.e.} Unit-1. The same procedure is followed to all other units. Here from data collected, assessed the exponentiality of the data tested via the Chi-square GoF test. Firstly descriptive statistics are needed to be calculated (Table 2). Using these, we can define the subintervals.

Table 1 shows Repair Hours of Unit 1 (Sorted Data).

| Table 1: Repair Hours of Unit 1 of Category-I. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0.02            | 0.05            | 0.06            | 0.08            | 0.08            | 0.08            | 0.08            |
| 0.09            | 0.10            | 0.11            | 0.13            | 0.13            | 0.14            | 0.14            |
| 0.14            | 0.15            | 0.15            | 0.15            | 0.16            | 0.16            | 0.16            |
| 0.17            | 0.18            | 0.19            | 0.20            | 0.20            | 0.21            | 0.22            |
| 0.22            | 0.22            | 0.23            | 0.24            | 0.28            | 0.30            | 0.31            |
| 0.32            | 0.38            | 0.47            | 0.48            | 0.49            |                |                |

The hypothesis of the problem is:

\( H_0 \): The data follows Exponential Distribution.
\( H_a \): The data do not follow the Exponential Distribution.
For the above data ($\lambda = 0.19$) is generated on exponential distribution assessment. Now, we will assess the exponentiality of the data via the Chi-square GoF test.

We obtain the descriptive statistics given in Table 2.

Table 2: Descriptive Statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>StdDev</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>40</td>
<td>0.19</td>
<td>0.16</td>
<td>0.11</td>
<td>0.02</td>
<td>0.49</td>
<td>0.13</td>
<td>0.22</td>
</tr>
</tbody>
</table>

For endpoints, we now select 0.10, 0.15, 0.20, 0.30 and 0.50, which in turn define subintervals. We obtain cumulative and individual cell probability values.

For ex,

$$P_{0.19}(0.10) = 1.0 - \exp\left[-\frac{0.10}{0.19}\right] = 0.41$$

The resulting values are shown in Table 3.

The result of the Chi-square GoF test statistic for this data is

$$\chi^2 = \sum \left( \frac{O_i - E_i}{E_i} \right)^2 = 11.33$$

The critical value at 5% level of significance and 3 df is 7.81. Since tabulated value is greater than the critical value, we accept $H_0$ and conclude the Exponential distribution is a good fit to the data.

Similarly, we calculate Chi-square Test of GoF of Unit I of Category II. The Repair of Hours of Unit I of Category II is shown in Table 4.

For the data given in Table 4 ($\lambda = 6.39$) is generated on exponential distribution assessment. Now, we will assess the exponentiality of the data via the Chi-square Goodness of Fit test. We obtain the descriptive statistics given in Table 5.

Like how we did earlier, for endpoints we now select 0.6, 1.5, 2.5, 6.5, 9, 20 and 25, which in turn define subintervals. We obtain cumulative and individual cell probability values. For Ex,

$$P_{6.39}(0.6) = 1.0 - \exp\left[-\frac{0.6}{6.39}\right] = 0.09$$

The resulting values are shown in Table 6.

The result of the Chi-square GoF test statistic for this data is

$$\chi^2 = \sum \left( \frac{O_i - E_i}{E_i} \right)^2 = 4.98$$

The critical value at 5% level of significance and 5 df is 11.07. Since, tabulated value is less than the critical value, we reject $H_0$ and conclude that the exponential distribution is not a good fit to the data.

The procedure explained above to obtain goodness of fit is for the Unit-I. Similar type of analysis can be applied to other units also. Results obtained for other units along with Unit-I are given in Table 7 along with their chi-square value to test the GoF.

6. Analysis

Tables 2, 3, 5, 6 and 7 gives the result obtained for estimating the statistical parameter for Repair Hours Category-I and Category-II of Unit-I Repair Time data are fitted with exponential distribution. The applicability
Estimation of Reliabilities of Thermal Power Plant – A Case Study

One of the very important conclusions is that fitting of repair hours for the data related to power sector necessarily requires data to be spitted into two categories as minor repair hours and major repair hours.

2. For major repair hours, the data was found to be good fit with exponentiality.

3. For minor repairs, further investigation is required to fit the data.

It implies the implementation of exponential distribution to the study concludes that the Failure Rate is not constant but decreases with time.

Remarks

The above procedure can also be extended to show that the Weibull and Lognormal distributions also give a good fit for the observed failure and working times of power units.

### Foot Note
No Star: Accepted at both α = 5% and α = 1%, * : Rejected at 5%, **: Rejected at 1%.
C. P. S. Hungund and Shrikant Patil

Acknowledgement

Authors sincere thanks to Dr. K.L.A.P. Sharma, Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapur; for his useful suggestion and valuable comments in the presentation of the paper.

References


RELIABILITY ANALYSIS OF THERMAL POWER GENERATING UNITS BASED ON WORKING HOURS

C.P.S. Hungund and Shrikant Patil
Department of Statistics, Gulbarga University, Gulbarga
E Mail: cpshgu@gmail.com, shrikant07.inc@gmail.com

Received February 07, 2014
Modified April 25, 2014
Accepted May 17, 2014

Abstract
For any system reliability calculation, estimation plays a vital role and hence the analysis is to be diverted to calculate reliabilities of the units. To do this, we need to analyze failure rates or inter-failure time intervals or duration of the failure free operation of the units. This data is usually termed as Working Hours (successive failures). Working Hours implies that the gap between failure to failure which is directly linked with the reliability of the system. Hence in this paper, first we fit and test the suitability of the data using Exponential and Weibull distribution through chi-square Goodness of Fit (GoF) test. In the later part, reliabilities of different units are calculated using the same distributions. Conclusions are drawn based on the results obtained.

Key Words: Repair Hours, Failure Rate, Exponential Model, Weibull Model, Reliability, Working Hours.

1. Introduction
Thermal power generation being one of the important power sources among other power sources, now days exposed to lots of challenges. System complexity, poor design, expertise human resource capital, lack of maintenance, climatic conditions, scarcity of materials and so on are some of the causes of failure that are gaining more importance. Due to all these issues, generation of electricity has been suffering from long time. The facts and figures stated by Ministry of Power further tell the seriousness of the problem; among them, some are given below:

- Power sector is one of the fastest growing sectors in India, which essentially supports the economic growth. The power sector needs to grow at the rate of 12% to maintain the present GDP growth of 8%.
- Presently the energy deficit is about 8.3% and the power shortage during the peak period is about 12.5%.
- The total installed capacity of the power generating units is about 1,24,310 MW. Thermal power generating units contribute 66.4% of total installed capacity. The average plant load factor of the thermal power generating units is 74.8%.
  (Source: Ministry of power website – www.powermin.nic.in)

Thus, in the present scenario, apart from capacity escalation, there is an immense need to improve the performance of the individual thermal power generating units. The performance improvement of individual thermal power generating units will
help in achieving increased power generation, reduction in power generation cost and thereby reducing the demand and improving the competitiveness of the Indian power industry.

A probabilistic analysis of the system under given operative conditions is helpful in understanding the behavior of the units of thermal power plant which further helps in statistical modeling to analyze failure in the system i.e. to optimize the system working. Therefore in order to correctly assess the problem and efficient functioning of the thermal units, it is very essential to correctly model and test it through probabilistic approach. Thus, the present paper investigates the suitability of Exponential and Weibull distributions through chi-square test of GoF regarding failures of thermal power unit data. Since higher availability of a power plant is depending upon higher reliability, the study is extended to reliability analysis to analyze the reliability of each unit of thermal power plant. For the present study, we have chosen Raichur Thermal Power Station (RTPS), Raichur, Karnataka to study thermal units. A unit represents the system consisting of stem generator, boiler furnace and stems drum, super heater, reheater, stem turbine, auxiliary systems, fuel preparation system etc involving in electricity production. There are seven such power generating units at RTPS. Working hours and Repair hours are collected from the seven units during the period 2004-2011.

2. Survey of Literature

Adhikary et al. [1] dealt a case of “RAM investigation of coal-fired thermal power plants: A case study”. The authors have investigated the reliability, availability and maintainability (RAM) characteristics of a 210 MW coal-fired thermal power plant (Unit-2) from a thermal power station in eastern region of India. Analyses of components/equipments have been tested through distribution, later GoF test have been performed through Kolmogorov–Smirnov Test. Critical mechanical subsystems with respect to failure frequency, reliability and maintainability are identified for taking necessary measures for enhancing availability of the power plant and the results are compared with Unit-1 of the same Power Station. The author concludes that RAM analysis is very much effective in finding critical subsystems and deciding their preventive maintenance program for improving availability of the power plant as well as the power supply.

Fernando Jesus Guevara Carazas and Gilberto Francisco Martha de Souza [2] consider the case of “Availability Analysis of Gas Turbines Used in Power Plants”. In the paper, the reliability analysis is performed for two gas turbines installed in the power plant. The reliability analysis is based on the time to failure data analysis. The method allows one to carry out system reliability, maintainability and availability analyses. Reliability tests have been performed through Exponential, Weibull and Lognormal Distributions. Both gas turbines’ reliability and availability estimates were considered as preliminary. The improvement of ‘time to failure’ and ‘time to repair’ databases during future operational years (with the addition of more failure and repair data) allows more reliable estimates of the turbines reliability and availability. Nevertheless these estimates can be used to check design and maintenance procedures in order to adopt them to the gas turbine local operational condition that may be different from the average condition considered in the equipment design. The author concludes that, these estimates can also be used for benchmarking in order to compare the performance of the same gas turbine model operating in different sites.
Krishna Reddy [6] dealt a case of “Modeling the Causes of Production Downtimes: An Empirical Study of a Thermal Power Generating Unit”. An attempt is made to find which distribution is best fitted for working time between successive failures (working hours) and downtime (repair hours). Exponential and Weibull distributions are commonly assumed model of failure time and repair time. Chi-square test is used to test the GoF.

Romeu [8] discussed GoF test using various distributions in “The Chi-Square: a Large Sample Goodness of Fit Test”. In the study, the author has developed a model to fit the data using Normal Distribution, Exponential Distribution and Weibull Distribution. Then Chi-Square test was applied to know the underlying distribution supports the data or not. The author has taken several examples namely Normal, Lognormal, Exponential and Weibull distributions and showed how the GoF can be dealt via Chi-Square test.

Hungund and Patil [4] analyzed and compared critically using repair hours of different power units using Exponential distribution. It has been observed that the distribution did not give a good fit to the repair hour data for all the power generating units. Hence it was concluded that the procedure for fitting the data need to be further developed.

3. Objectives
The objectives of the present study are:
- To fit the Exponential and Weibull distribution for working hours of seven power units of a thermal power plant.
- To test the suitability of Exponential and Weibull through the chi-square test of GoF.
- To test the reliability of seven power units of a thermal power plant.
- To compare the Exponential and Weibull reliabilities.

4. Goodness of Fit Test
As it was stated that the failures of thermal units are due to various reasons. In order to identify the best suitable analysis, the collected data has been split into two categories viz., Category-I and Category-II. Category-I is based on minor working hours and category-II is based on major working hours.

4.1 Fitting of Exponential Distribution
Exponential distribution is commonly assumed model for time to failure data. Here an effort is made to fit the data using Exponential distribution for failures (working hours).

4.1.1 Probability Density Function (PDF)
The PDF of an Exponential distribution is

\[ f(x; \lambda) = \lambda e^{-\lambda x}, \ x \geq 0 \]  

where \( \lambda \) is the parameter.

The parameters are usually unknown and are to be estimated from the data. It is also necessary to check whether the distribution of observed data fits into one of the
known theoretical distributions or not, so that the distribution can be better understood if necessary.

4.1.2 Cumulative Distribution Function (CDF)

The CDF of Exponential distribution is given by:

\[ F(x; \lambda) = 1 - e^{-\lambda x}, \quad x \geq 0 \]  

(4.2)

4.1.3. Testing of Hypothesis

The hypothesis of the problem is given by:

\[ H_0: \text{Exponential Distribution is a good fit for the data} \]

\[ H_a: \text{Exponential Distribution is not a good fit for the data} \]

Using the procedure developed in the paper Hungund and Patil [4], we obtain the following results for fitting an Exponential distribution for whole and categorical data on working hours for seven units which are summarized in tables 1 and 2 with chi-square value to test its GoF.

<table>
<thead>
<tr>
<th>Units</th>
<th>n</th>
<th>k</th>
<th>Mean</th>
<th>Lamda</th>
<th>Chi-Square</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>74</td>
<td>10</td>
<td>26.55</td>
<td>0.038</td>
<td>29.85</td>
<td>6**</td>
</tr>
<tr>
<td>Unit 2</td>
<td>94</td>
<td>11</td>
<td>13.4</td>
<td>0.4</td>
<td>182.29</td>
<td>7**</td>
</tr>
<tr>
<td>Unit 3</td>
<td>56</td>
<td>7</td>
<td>38.78</td>
<td>0.026</td>
<td>11.83</td>
<td>3**</td>
</tr>
<tr>
<td>Unit 4</td>
<td>59</td>
<td>9</td>
<td>40.33</td>
<td>0.025</td>
<td>16.74</td>
<td>5**</td>
</tr>
<tr>
<td>Unit 5</td>
<td>52</td>
<td>8</td>
<td>44.05</td>
<td>0.023</td>
<td>7.45</td>
<td>5</td>
</tr>
<tr>
<td>Unit 6</td>
<td>46</td>
<td>6</td>
<td>51.7</td>
<td>0.019</td>
<td>1.81</td>
<td>3</td>
</tr>
<tr>
<td>Unit 7</td>
<td>79</td>
<td>8</td>
<td>27.51</td>
<td>0.036</td>
<td>13.14</td>
<td>5**</td>
</tr>
</tbody>
</table>

Table 1: Summary of Exponential distribution working hours for whole data

<table>
<thead>
<tr>
<th>Units</th>
<th>Category</th>
<th>n</th>
<th>k</th>
<th>Mean</th>
<th>Lamda</th>
<th>Chi-Square</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>Category-I</td>
<td>40</td>
<td>5</td>
<td>5.15</td>
<td>0.194</td>
<td>9.86</td>
<td>3*</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>34</td>
<td>5</td>
<td>51.72</td>
<td>0.019</td>
<td>5.76</td>
<td>1*</td>
</tr>
<tr>
<td>Unit 2</td>
<td>Category-I</td>
<td>50</td>
<td>6</td>
<td>1.46</td>
<td>0.685</td>
<td>10.55</td>
<td>3*</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>44</td>
<td>6</td>
<td>26.97</td>
<td>0.037</td>
<td>5.56</td>
<td>4</td>
</tr>
<tr>
<td>Unit 3</td>
<td>Category-I</td>
<td>28</td>
<td>4</td>
<td>5.36</td>
<td>0.187</td>
<td>3.1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>28</td>
<td>4</td>
<td>72.21</td>
<td>0.014</td>
<td>6</td>
<td>1*</td>
</tr>
<tr>
<td>Unit 4</td>
<td>Category-I</td>
<td>30</td>
<td>5</td>
<td>5.95</td>
<td>0.168</td>
<td>1.54</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>29</td>
<td>4</td>
<td>75.89</td>
<td>0.013</td>
<td>4.17</td>
<td>2</td>
</tr>
<tr>
<td>Unit 5</td>
<td>Category-I</td>
<td>26</td>
<td>4</td>
<td>8.78</td>
<td>0.114</td>
<td>0.37</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>26</td>
<td>4</td>
<td>79.31</td>
<td>0.013</td>
<td>1.18</td>
<td>2</td>
</tr>
<tr>
<td>Unit 6</td>
<td>Category-I</td>
<td>24</td>
<td>3</td>
<td>11.05</td>
<td>0.09</td>
<td>3.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>22</td>
<td>3</td>
<td>96.04</td>
<td>0.01</td>
<td>2.15</td>
<td>1</td>
</tr>
<tr>
<td>Unit 7</td>
<td>Category-I</td>
<td>40</td>
<td>4</td>
<td>4.91</td>
<td>0.204</td>
<td>0.75</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Category-II</td>
<td>39</td>
<td>4</td>
<td>50.64</td>
<td>0.02</td>
<td>4.56</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2: Summary of Exponential distribution working hours data for categorical data.

*: Rejected at 5%, **: Rejected at 5% & 1%.
The analysis of the results obtained in Tables 1 & 2 estimating the statistical parameter for working hours of seven units imply that the working time data are fitted with Exponential and Weibull distributions. From Table 1, it is inferred that the Exponential distribution is not a good fit to the data on working hours, except for unit 5 and unit 6 at 5% level of significance and unit 5, 6, 7 at 1% level of significance. Hence we split the data into two categories. From table 2 it can be observed that the Exponential distribution gave a good fit for categorical data of working hours for all the units at 1% level of significance. At 5%, except unit-1 and category-2 of unit-3, the data of all other units show Exponential distribution gave a good fit for working hours. We claim that splitting of data into two categories minor repair hours and major repair hours helps to analyze failure in the system i.e. to optimize the working hours in the system.

4.2 Fitting of Weibull Distribution

Romeu [9] discusses some empirical and practical methods for checking and verifying the statistical assumptions of the Weibull distribution in the paper “Empirical Assessment of Weibull Distribution”. The author has elaborated the statistical assumptions of Weibull distribution by taking several numerical and graphical examples. Two approaches were used to assess the distribution assumptions. One is by implementing numerically convoluted, theoretical GoF and other practical procedures that are based on intuition and graphical distribution properties. In the following sections an attempt is made to show that the data assumes Weibull distribution by verifying the statistical assumptions of Weibull distribution numerically and graphically.

4.2.1 The PDF and CDF of Weibull Distribution

Consider the PDF and CDF of Weibull distribution

\[
f(x) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\} \tag{4.3}
\]

\[
F(x) = 1 - \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\} \tag{4.4}
\]

4.2.2 Estimation of Weibull Parameters

Taking the logarithms of the distribution function \(F(x)\) (eqn. 4.4) and doing some algebra, we obtain:

\[
\text{Log} \left\{ \text{Log} \left( \frac{1}{1 - F(x)} \right) \right\} = -\beta \text{Log}(\alpha) + \beta \text{Log}(x) \tag{4.5}
\]

When the distribution of the failure is really Weibull, the above equation is that of a line. Now assume that an estimation of \(F(x)\) can be obtained and denote it \(P_x\). We then can substitute \(P_x\) in lieu of \(F(x)\) in the equation and solve for \(x\). To estimate \(P_x\) for any data point \(x\), i.e. the median rank by defining:

\[
F(x) \approx P_x = \frac{\text{Rank}(x) - 0.3}{n + 0.4} \tag{4.6}
\]
Where Rank(x) is the rank of x, in the sorted sample. Using such Px values, we can plot the regression line on Log(Log(1/(1 – Px))) call it as Y against Log_e(X) call it as X which is shown in the figure-1.

The collected data (Table 3) has been used to develop a model to fit the data of power units and tested good fit to the working hours for all seven units i.e. Unit-1 to Unit-7.

<table>
<thead>
<tr>
<th>x</th>
<th>LogY</th>
<th>LogX</th>
<th>Log(1/(1 – Px))</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.14</td>
<td>0.15</td>
<td>0.29</td>
<td>0.57</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.15</td>
</tr>
<tr>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.08</td>
</tr>
<tr>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.07</td>
<td>5.00</td>
<td>6.00</td>
</tr>
<tr>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.05</td>
</tr>
<tr>
<td>12</td>
<td>13.00</td>
<td>13.00</td>
<td>14.81</td>
<td>16.00</td>
<td>17.00</td>
</tr>
<tr>
<td>18.87</td>
<td>19.00</td>
<td>20.38</td>
<td>20.40</td>
<td>21.00</td>
<td>23.00</td>
</tr>
<tr>
<td>24.00</td>
<td>25.00</td>
<td>26.00</td>
<td>27.90</td>
<td>27.93</td>
<td>30.14</td>
</tr>
<tr>
<td>31.00</td>
<td>32.00</td>
<td>45.00</td>
<td>45.71</td>
<td>46.97</td>
<td>46.97</td>
</tr>
<tr>
<td>49.81</td>
<td>68.26</td>
<td>69.32</td>
<td>72.00</td>
<td>76.00</td>
<td>80.00</td>
</tr>
<tr>
<td>109.00</td>
<td>119.45</td>
<td>162.44</td>
<td>198.91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Working Hours of Unit 1

We obtain the intercept and slope of regression equation for the data X and Y using Microsoft Excel function ‘=intercept(Y,X)’ and ‘=slope(Y,X)’ respectively. Thus we have intercept=-1.9118 and slope (β)=0.6385. The characteristics life (α) of Weibull can be obtained computing

\[ \text{CharacteristicLife (}\alpha) = \exp \left[ -\left( \frac{\text{Intercept}}{\text{Slope}} \right) \right] = \exp \left[ -\left( \frac{-1.9118}{0.6385} \right) \right] = 19.9772 \]

Then the estimated regression equation will be \( Y = -\beta \log(\alpha) + \beta x \).

The following table summarizes the regression results.

<table>
<thead>
<tr>
<th>Intercept - ( \beta \log(\alpha) )</th>
<th>Slope (( \beta ))</th>
<th>CharLf (( \alpha ))</th>
<th>Correlation (( r ))</th>
<th>Coeff of Determination (( r^2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>-1.9118</td>
<td>0.6385</td>
<td>19.9772</td>
<td>0.9951</td>
</tr>
</tbody>
</table>

Table 4: Summary of Regression equation.
From the figure-1 it is evident that the data is Weibull, since the plot of the transformed data is linear with slope $\beta$ and intercept $-\beta \log(\alpha)$.

**4.2.3 Testing of Hypothesis**

The parameter estimation allows us to define the composite distribution hypothesis

$H_0$: The distribution of the population originating the data set is Weibull.

$H_a$: The distribution of the population originating the data set is not Weibull.

The above hypothesis is tested via the Chi-Square GoF test. For endpoints we now select 1.00, 3.00, 8.00, 12.00, 17.00, 23.00, 31.00, 48.00, 85.00 and 200.00 which in turn define subintervals. We also obtain the cumulative and individual cell probability values. For the first point, 1.0:

$$P_{\alpha=1.00} = 1 - \exp\left\{-\left(\frac{1.0}{X_\alpha}\right)^\beta\right\} = 1 - \exp\left\{-\left(\frac{1.0}{19.9772}\right)^{0.6385}\right\} = 0.14$$

The resulting values are shown in the following table 5.

<table>
<thead>
<tr>
<th>X</th>
<th>CumProb</th>
<th>CellProb</th>
<th>Expected</th>
<th>Observed</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14</td>
<td>0.14</td>
<td>10.17</td>
<td>12</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>0.12</td>
<td>8.91</td>
<td>8</td>
<td>0.09</td>
</tr>
<tr>
<td>8</td>
<td>0.43</td>
<td>0.17</td>
<td>12.55</td>
<td>8</td>
<td>1.65</td>
</tr>
<tr>
<td>12</td>
<td>0.51</td>
<td>0.09</td>
<td>6.44</td>
<td>8</td>
<td>0.38</td>
</tr>
<tr>
<td>17</td>
<td>0.59</td>
<td>0.08</td>
<td>5.92</td>
<td>6</td>
<td>0.00</td>
</tr>
<tr>
<td>23</td>
<td>0.67</td>
<td>0.07</td>
<td>5.25</td>
<td>7</td>
<td>0.59</td>
</tr>
<tr>
<td>31</td>
<td>0.73</td>
<td>0.07</td>
<td>5.09</td>
<td>8</td>
<td>1.67</td>
</tr>
<tr>
<td>48</td>
<td>0.83</td>
<td>0.09</td>
<td>6.83</td>
<td>6</td>
<td>0.14</td>
</tr>
<tr>
<td>85</td>
<td>0.92</td>
<td>0.09</td>
<td>6.91</td>
<td>6</td>
<td>0.12</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>0.08</td>
<td>5.95</td>
<td>5</td>
<td>0.15</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>74</td>
<td>74</td>
<td>74</td>
<td>5.08</td>
</tr>
</tbody>
</table>

**Table 5: Intermediate Values for the Weibull distribution GoF test.**
Since we have estimated both $\alpha$ and $\beta$ the resulting chi-square has degrees of freedom $(df) = k-2-1=10-2-1=7$. Here the chi-square statistic value $(5.08)$ is not larger than the chi-square table value $14.03$ for $7$ df and $\alpha=0.05$. Therefore we accept $H_0$ and conclude that the Weibull distribution is good fit to the data.

The procedure explained above to obtain GoF is for the Unit-1. Similar procedure can be applied to other units also. Results obtained for other units along with Unit-1 are given in Table 6 with their chi-square values to test the GoF.

<table>
<thead>
<tr>
<th>Units</th>
<th>n</th>
<th>k</th>
<th>Mean</th>
<th>Chi-Square</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>74</td>
<td>10</td>
<td>26.55</td>
<td>5.08</td>
<td>7</td>
</tr>
<tr>
<td>Unit 2</td>
<td>94</td>
<td>11</td>
<td>13.4</td>
<td>4.32</td>
<td>9</td>
</tr>
<tr>
<td>Unit 3</td>
<td>56</td>
<td>7</td>
<td>38.78</td>
<td>4.10</td>
<td>3</td>
</tr>
<tr>
<td>Unit 4</td>
<td>59</td>
<td>9</td>
<td>40.33</td>
<td>4.20</td>
<td>5</td>
</tr>
<tr>
<td>Unit 5</td>
<td>52</td>
<td>8</td>
<td>44.05</td>
<td>7.45</td>
<td>4</td>
</tr>
<tr>
<td>Unit 6</td>
<td>46</td>
<td>6</td>
<td>51.7</td>
<td>3.84</td>
<td>3</td>
</tr>
<tr>
<td>Unit 7</td>
<td>79</td>
<td>8</td>
<td>27.51</td>
<td>1.22</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 6: Intermediate Values for the Weibull distribution GoF test.

It is interesting to note that the collected data on working hours need not be split into categories in case of Weibull distribution (Table 6). The analysis of the results obtained in Table 6 infer that the model gave a good fit to working hours for the entire unit both at $1\%$ and $5\%$ level of significance for the period of seven years.

5. Reliability Analysis

Reliability can be defined as the probability that a system will perform properly for a specified period of time under a given set of operating conditions. Implied in this definition is a clear-cut criterion for failure, from which one may judge at what point the system is no longer functioning properly. For the thermal power units the failure criterion is any component failure that causes incapacity of generating the nominal power output. The reliability analysis is performed for each of the thermal units in the power plant. The reliability analysis is based on the time to failure data analysis. To better understand the behavior of lifetime distributions, reliabilities of Exponential and Weibull analysis is performed in the following sections.

5.1 Reliability of Exponential Distribution

The Exponential distribution is one of the most widely used in reliability problems (particular case of a Weibull distribution with the shape parameter $\beta = 1$). The very reason behind popularity of this distribution can be attributed primarily to the fact that it is be used to model the time to failure of components and systems with constant failure rate, a situation that is often realistic. The reliability function is given below:

$$R(t) = e^{-\lambda t}$$

(5.1)

The mean time to failure of the Exponential function is simply the inverse of the failure rate $\lambda$. 
Reliability Analysis of Thermal Power Generating Units Based on...

\[ MTTF = \frac{1}{\lambda} \]  \hspace{1cm} (5.2)

5.2 Reliability of Weibull distribution

The two-parameter Weibull distribution is used to model failures is represented by the following equation:

\[ R(t) = \exp \left( -\left( \frac{t}{\alpha} \right)^\beta \right) \]

Where \( R(t) \) reliability at time \( t \), \( t \) time period \([h]\), \( \beta \) is Weibull distribution shape parameter and \( \alpha \) is Weibull distribution characteristic life \([h]\).

5.3 Reliability analysis of Exponential and Weibull distribution

The Weibull distribution parameters are estimated through the use of regression line that fit the distribution to the ‘time to failure’ data. The following table shows the Weibull parameters for all the units.

<table>
<thead>
<tr>
<th></th>
<th>Unit1</th>
<th>Unit2</th>
<th>Unit3</th>
<th>Unit4</th>
<th>Unit5</th>
<th>Unit6</th>
<th>Unit7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>26.56</td>
<td>13.4</td>
<td>38.78</td>
<td>40.33</td>
<td>44.05</td>
<td>51.7</td>
<td>27.51</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.91182</td>
<td>-1.05555</td>
<td>-2.11772</td>
<td>-2.32664</td>
<td>-3.05913</td>
<td>-2.63672</td>
<td>-2.27999</td>
</tr>
<tr>
<td>Slope (( \beta ))</td>
<td>0.638477</td>
<td>0.511687</td>
<td>0.626185</td>
<td>0.679697</td>
<td>0.835805</td>
<td>0.695358</td>
<td>0.731548</td>
</tr>
<tr>
<td>Alpha (( \alpha ))</td>
<td>19.97222</td>
<td>7.868664</td>
<td>29.42774</td>
<td>30.66288</td>
<td>38.86505</td>
<td>44.34036</td>
<td>22.57102</td>
</tr>
</tbody>
</table>

Table 7: Summary of regression estimators for all the units.

Depending upon the values of the shape parameter (\( \beta \)), Weibull failure rate increases (\( \beta>1 \)), decreases (\( \beta<1 \)) or remains constant (\( \beta=1 \)). Table 7 infers that all the units have reliability distribution with shape parameter less than 1. When 0<\( \beta<1 \), the distribution has decreasing failure rate. The reliability distribution for Exponential and Weibull distribution curve for all the units are presented in table 8 as well as figure-2. The points presented in the graph represent the median rank plotting reliability estimate for each of the time to failure data, arranged in increasing order. Those points are used to verify the adherence of the reliability distribution to the failure data.
Table 8: Reliability of the seven units of Thermal Power Plant at different time units ‘t’.

<table>
<thead>
<tr>
<th>Time 't'</th>
<th>Unit1</th>
<th>Unit2</th>
<th>Unit3</th>
<th>Unit4</th>
<th>Unit5</th>
<th>Unit6</th>
<th>Unit7</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.83</td>
<td>0.71</td>
<td>0.69</td>
<td>0.55</td>
<td>0.83</td>
<td>0.76</td>
<td>0.83</td>
</tr>
<tr>
<td>10</td>
<td>0.69</td>
<td>0.59</td>
<td>0.47</td>
<td>0.42</td>
<td>0.77</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>15</td>
<td>0.57</td>
<td>0.50</td>
<td>0.33</td>
<td>0.35</td>
<td>0.68</td>
<td>0.58</td>
<td>0.69</td>
</tr>
<tr>
<td>20</td>
<td>0.47</td>
<td>0.43</td>
<td>0.22</td>
<td>0.29</td>
<td>0.60</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>25</td>
<td>0.39</td>
<td>0.38</td>
<td>0.15</td>
<td>0.25</td>
<td>0.52</td>
<td>0.47</td>
<td>0.54</td>
</tr>
<tr>
<td>50</td>
<td>0.15</td>
<td>0.22</td>
<td>0.02</td>
<td>0.14</td>
<td>0.28</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>100</td>
<td>0.02</td>
<td>0.10</td>
<td>0.00</td>
<td>0.06</td>
<td>0.08</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>150</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>300</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Figure-2: Graphs of the reliability of seven units at different time units t.

From Table 8, it is evident that Unit-6 in Exponential and Unit-5 in Weibull distribution shows higher reliability. This indicates that Unit-5 and Unit-6 are getting close to the period of random failures characterized by the shape parameter equal to 1. Whereas Unit-2 is showing lesser reliability in both the distribution when compared with other units which indicates necessary measure has to be taken for improvement. From figure-2 in the Weibull plot, though Unit-6 is showing lesser reliability initially compared to unit-5, but after 25th hours, Unit-6 starts showing good reliability than Unit-5. This concludes that unit-6 is best performing unit among the other six units. Also note that reliability of power units after 100 hours, Exponential reliability is below 0.10 mark, whereas its counterpart i.e. Weibull, still having better reliabilities. Even if we carefully observe, till 50th hour, both reliabilities are more or less same. After that Exponential reliabilities start deviating (decreasing) with Weibull reliabilities, and it continues to be so for subsequent hours. The very reason we believe that Weibull distribution found good fit to the data for whole data.

6. Conclusion

The reliability analysis of thermal power generating units based on working hours have been tested through Exponential and Weibull distributions. Later GoF have been performed through chi-square test. The GoF test reveals that Weibull distribution
is the most reliable distribution for fitting working hour data. Reliabilities are identified for taking necessary measures enhancing availability of the power plant.

Acknowledgement
Authors express sincere thanks to Dr. K. L. A. P. Sarma, (Retd.) Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapur for his useful suggestions and valuable comments in the presentation of this paper.

References
APPENDIX
REFERENCES


