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)

-1.9741	-1.9301	-1.2274	-0.5704	0.0000	0.0000	0.0000
0.0005	0.0006	0.0006	0.0007	0.0008	0.0079	0.4055
0.6931	0.6931	0.6931	1.0986	1.0986	1.1244	1.3863
1.3863	1.3863	1.3863	1.4035	1.6094	1.7918	2.0794
2.3026	2.3026	2.3026	2.3026	2.3026	2.3071	2.4849
2.4849	2.5649	2.5649	2.6956	2.7726	2.8332	2.8332
2.9375	2.9444	3.0145	3.0156	3.0445	3.1355	3.1355
3.1781	3.2189	3.2581	3.3286	3.3296	3.4057	3.4323
3.4340	3.4657	3.8067	3.8223	3.8495	3.8496	3.8775
3.9083	4.2234	4.2387	4.2767	4.3307	4.3820	4.4530
4.6913	4.7829	5.0903	5.2929			

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

Statistics	**N**	**Mean**	**Median**	**Std. Dev**	**Q1**	**Q3**
Value	74	2.2669	2.5649	1.6731	1.1051	3.4257

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: $(\mu-\sigma, \mu+\sigma)$ =(0.5938,3.9400) includes 50 values representing 67.57% of the total data set close to the expected 68.25%. The interval $(\mu-2\sigma, \mu+2\sigma)$ = (-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 $\mu\pm3\sigma$, supporting the statement that about 1% of the values would be outside the interval $(\mu-3\sigma, \mu+3\sigma)$. 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.

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

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

$$P_{2.2687,1.6709} = Normal\left(\frac{-1.9741-2.2669}{1.6731}\right) = Normal(-2.5348) = 0.0056$$

Similarly, Pi 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:

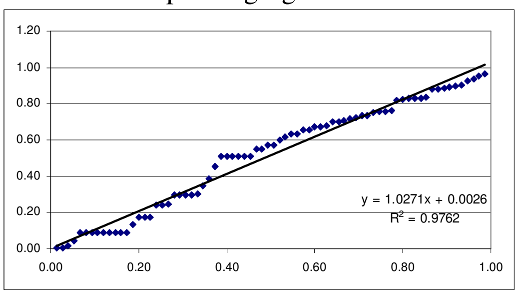

Figure 1: Normal Probability Plot

When the population is normal, the probability plot shown in above figure-1 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 μ and standard deviation σ where μ>0 and σ>0. If the probability density function of X is:

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{\frac{1}{2\sigma^2}(\ln(x)-\mu)^2} \text{ for x>0.}$$

5.1.2 Cumulative Distribution Function

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

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

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: The data follows lognormal distribution.

Ha: 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 μ=2.2669 & σ=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

Ex $P_{2.2669,1.6731} = Normal\left(\frac{0.0-2.2669}{1.6731}\right) = Normal(-1.3549) = 0.0877$

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

Table 3: Intermediate Values for the lognormal distribution GoF test

						Pooled Data		
X	**StdEnd**	**CumProb**	**CellProb**	**Ei**	**Oi**	**Ei**	**Oi**	**Chi-Sq**
0	-1.3549	0.0877	0.0874	6.4676	7	9.8118	14	1.7877
0.4055	-1.1126	0.1329	0.0452	3.3465	7	12.3175	10	0.436
1.3863	-0.5263	0.2993	0.1664	12.3119	10	15.4978	9	2.7244
2.3026	0.0213	0.5085	0.2092	15.4788	9	9.187	9	0.0038

2.8332	0.3384	0.6325	0.124	9.1752	9
3.1355	0.5191	0.6982	0.0657	4.8598	7
3.434	0.6975	0.7573	0.0591	4.3735	8
3.8712	0.9588	0.8312	0.0739	5.4699	6
4.4427	1.3004	0.9033	0.0721	5.3343	6
5.2983	1.8118	1	0.0967	7.1585	5
Total			***1.00***	***74.00***	***74***

4.8651	7	0.9369
9.8493	14	1.7492
5.3325	6	0.0836
7.1364	5	0.6396
74.00	***74***	***8.3611***

The result of the Chi-Square Goodnes of Fit test statistic for this data is: $\chi^2 = \sum \frac{(Oi - Ei)^2}{Ei} = 8.3611$

The critical value at 5% level of significance and 5 df is 11.07. Since tabulated value is less than the critical value, we fail to reject H_0 and conclude that the lognormal distribution is 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 along with Unit-1 are given in Table No. 4.

Table 4: Summary of the Lognormal distribution for working Hours.

Units	N	k	Mean	StdDev	Chi-Square	DF
Unit 1	74	10	2.2669	1.6731	8.361	5
Unit 2	94	11	1.0834	2.2337	10.935	8
Unit 3	56	7	2.3908	2.0820	6.135	4
Unit 4	59	9	2.6129	1.8578	5.148	6
Unit 5	52	8	2.9757	1.3904	1.752	3
Unit 6	46	6	2.9932	1.6649	6.453	3
Unit 7	79	8	2.3690	1.6298	3.538	5

5.2 Fitting of Normal Distribution

The exponential, 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 functions of these distributions are frequently 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 μ and σ, 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

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

5.2.3. Testing of Hypothesis

The hypothesis of the problem is given by:

H_0: The data follows 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 Normall distribution for working Hours.

Units	n	k	Mean	Lamda	Chi-Square	DF
Unit 1	74	10	26.550	0.038	65.599	7
Unit 2	94	11	13.400	0.075	97.478	5
Unit 3	56	7	38.780	0.026	108.806	2
Unit 4	59	9	40.330	0.025	57.817	2
Unit 5	52	8	44.050	0.023	57.450	2
Unit 6	46	6	51.700	0.019	22.840	2
Unit 7	79	8	27.510	0.036	62.490	3

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 terms 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. 6 as well as figure-2. 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

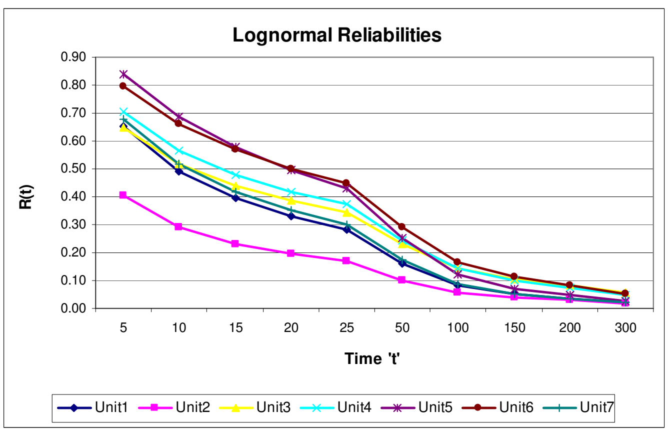

Figure 2: Graph shows the reliability of seven units at different time unit t.

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

1. Bollen,M.H.J. Literature Search for Reliability Data of Components in Electric Distribution Networks, Eindhoven University of Technology, 1993.
2. D.D. Adhikary *et al*. (2012). RAM investigation of coal-fired thermal power plants: A case study. The International Journal of Industrial Engineering Computations, Volume 3, pp. 423-434.
3. Hungund CPS and Shrikant Patil (2013) "Analysis of Production Downtimes in Thermal Power Plant using Exponential Failure Model – A Case Study", Bulletin for Pure and Applied Sciences, International Journal for Statistics and Mathematics, Volume 32E, Issue (No.1) 2013, Pp 113-120.
4. Hungund CPS and Shrikant Patil (2013) "Reliability Analysis of Thermal Power Generating Units based on Working hours". Paper is communicated.
5. Hungund CPS and Shrikant Patil (2013) "Estimation of Reliabilities of Thermal Power Plant – A Case Study", Paper is communicated.
6. J.M. Freeman, (1996),"Analysing equipment failure rates", International Journal of Quality & Reliability Management, Vol. 13 Iss: 4 pp. 39 – 49.
7. Kohichi Nakamura (1981). Reliability analysis of thermal power units based on the annual operating results Published online. Electrical Engineering in Japan, Volume 101, issued 4, Pages 78-87.
8. Krishna Reddy, B. (2008). Modeling the causes of Production downtimes: An empirical study of a Thermal Power Generating Unit, The ICFAI Journal of Operations Management, The Icfai University Press, Hyderabad May 2008 , pp. 33-40.
9. Nair ,P. and and Vasakarla Jayavani, Manual on Best Practices in Indian Thermal Power Generating Units. The IUP Journal of Infrastructure, The Icfai University Press, December 2006, Hyderabad.
10. Romeu, J.L. (2004). The Chi-Square: A Large Sample Goodness of Fit Test, RAC START, Volume 10 (Number 4). Publication of Reliability Analysis Center.
11. Romeu, J.L. (2002). Empirical Assessment of Normal and Lognormal Distribution Assumptions. RAC START, Volume 9 (Number 6). Publication of Reliability Analysis Center.