

University of Rajshahi

Rajshahi-6205

Bangladesh.

RUCL Institutional Repository

<http://rulrepository.ru.ac.bd>

---

Department of Statistics

MPhil Thesis

---

1991

# Statistical Analysis of the Average Life of Electric Bulbs: A Comparative Study

Islam, Md. Aminul

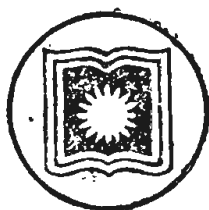
University of Rajshahi

---

<http://rulrepository.ru.ac.bd/handle/123456789/921>

*Copyright to the University of Rajshahi. All rights reserved. Downloaded from RUCL Institutional Repository.*

***STATISTICAL ANALYSIS OF THE AVERAGE  
LIFE OF ELECTRIC BULBS :  
A COMPARATIVE STUDY***



**THESIS SUBMITTED TO THE UNIVERSITY OF RAJSHAHI  
FOR THE DEGREE OF  
MASTER OF PHILOSOPHY**

**MD. AMINUL ISLAM**

**DEPARTMENT OF STATISTICS  
UNIVERSITY OF RAJSHAHI  
RAJSHAHI, BANGLADESH.**

**FEB, 1991**

## DECLARATION

This thesis contains no materials which have been accepted for the award of any other degree or diploma in any university. I hereby declare that the entire work now submitted as a thesis towards the fulfilment for the Master of Philosophy at the University of Rajshahi, Bangladesh, is the results of my own investigation.

*Mian M. A. Basher*  
(Md. Abul Basher Mian) *6/2/1991*  
Supervisor

**Associate Professor,  
Department of Statistics  
Rajshahi University.**

*Md. Aminul Islam*  
(Md. Aminul Islam)

ACKNOWLEDGEMENT

I wish to express my deepest sense of gratitude to my supervisor Mr. Abul Bashar Miah, Associate Professor, Department of Statistics, for his supervision, constant guidance, encouragement and helpful discussions throughout the course of this work.

I am also grateful to Professor K.M. Hossain, Professor Hosneara Hossain, Professor M.A. Sattar, Mr. M.A. Razzaque (Chairman, Department of Statistics), Mr. Korban Ali, Dr. Golam Mostafa, Dr. Samad Abedin, Dr. Quazi Abdus Samad, Dr. S.K. Bhattacharya for their valuable suggestions and helpful discussions.

Thanks are also due to Professor Sayeedur Rahman Khan, Dr. M.A. Sobhan and Dr. Akhtaruzzaman, Department of Applied Physics and Electronics, Rajshahi University for their help in the experimental part of the work. I would like to thank Mr. Fazlul Haque and Mr. Masudur Rahman, Instrument Engineer, Science Workshop, Rajshahi University for their assistance and cooperation.

I must express my gratitude to Professor A.K. Basak, Director, Rajshahi University Computer Centre for extending me the computer facilities. I am also thankful to Professor S.M. Murshed and Mrs. Shamima Sultana for analyzing the data for me. My thanks are also due to Professor M.A. Hamid, Department of Economics, Rajshahi University.

I wish to thank the following persons for their sympathetic treatment during the course of the work: Mr. A.H.M. Abul Kashem, Deputy Commissioner, Netrokona, Mr. Harun-ur-Rashid, Deputy Commissioner, Narayanganj,

Mr. Fakhrul Islam, A.D.C. (Gen), Netrokona, Mr. Bimal Kumar Kundu, A.D.C. (Rev), Netrokona, Mr. Patowary Nesar Uddin Ahmed, A.D.M., Netrokona.

I am indebted to the University of Rajshahi for awarding me a fellowship to pursue the research work.

The officers and staff of the Statistics Department also deserve my thanks for their cordial love. I wish to thank all the Research Fellows, residing at the Institute of Bangladesh Studies Hostel for their companionship and advice. I am indebted to my uncle Mr. Abdul Mannan, Deputy Controller, Rajshahi University, Mr. Abdus Samad Didar, Teacher, BIT, Rajshahi and their family who gave me affectionate love as their son. Mr. Ajit Kumar Chakrabarty also deserves my thanks for typing the thesis.

Finally, I would like to express my gratitude to my parents, sisters and brother for their inspiration and support to carry out this research work.

Md. Aminul Islam

SYNOPSIS

This thesis -- 'Statistical Analysis of the Average Life of Electric Bulbs: A Comparative Study' consists of six chapters of which chapter four and five are contributory, chapter six on the summary conclusion of the thesis. First three chapters of the thesis are introduction and discussions on the selection of a life testing model and the underlying methods of statistical inferences which has been used in the subsequent chapters towards the contribution in the thesis.

The introductory chapter one contains a statement of the problem we have undertaken a brief review of earlier studies and exploration of the possibilities of further work in relation to the present study, aims and scope of the study. It also provides a brief discussion on the concept of life testing and reliability and distributions of life times. A survey of some basic life testing models which has been used in contributory chapters has also been appended in this chapter.

Chapter two of this thesis is devoted on discussions on the selection of a life testing model to suit the analysis and prediction for a particular set of data. Almost all life testing data available for analysis and prediction are incomplete or censored. Most of the tools and techniques available in life testing literature for discriminating between competing life testing models are sensitive to the nature and size of censoring. This chapter contains a handful of

discussion on these problems and way out to face them. It also contains the methods of plotting procedures and some other parametric methods for discrimination among the family of models. A test of exponentiality against log-normal and a test of weibull against gamma model also appended in this chapter.

Chapter three of this thesis is a discussion on the methods of statistical inferences. As we have pointed out, most of the life testing data available for analysis are either incomplete or censored. Again censoring is not unique. There are different types of censoring and the form of likelihood changes with the type of censoring. Again a censored data available for analysis and prediction may be grouped or ungrouped. Methods of statistical inferences are sensitive to each and every characteristics of data and influences the prediction. Discussions on the use of methods of statistical inferences in relation to the characteristics of available data are the prime issues of this chapter. Methods of least-squares, minimum likelihood, censoring, Lifetable and Actuarial method, Product Limit estimator, Likelihood Ratio Test necessary for the subsequent contributory chapters are appended in this chapter.

For a comparative study among different brands of incandiscent electric bulbs, estimation of average life and other life parameters on actual life data are essential. Chapter four of this thesis is contributed on the estimation of reliability and life parameters. Data to suit our purpose were not readily available. In order to generate life data on incandiscent electric bulbs, we have done an experiment in the laboratory. Description of the experiment is appended at the beginning of

this chapter. It also contains screening and choice of models to suit the analysis of the data obtained from the experiment, least-squares and maximum likelihood estimates of life parameters, goodness of fit of the fitted model, and comparison of survival curves.

Selection of the best brand of incandescent electric bulb depends not only on the average life or reliability and life parameters but also on the intensity of light and consumption of power. Chapter five of this thesis is contributed towards the measurement and comparison of intensity of light and power consumption of different brands of incandescent electric bulbs.

This chapter contains experimental descriptions of measuring intensity of light and consumption of power of different brands of incandescent electric bulbs. It also provide voltage-intensity relationship. Estimates of intensity of light at average line voltage for different brands and power categories. A comparison of model component and overall intensity of light between brands of power categories are also appended in this chapter. Study on voltage-current relationship, prediction of flow of electricity through the filament of the bulbs in order to estimate the power consumption, estimates of power consumption at the average line voltage, comparison of component and overall power consumption between brands of power categories are also included in this chapter. It also provides estimates of cost-benefit factors and finally cost per unit of light for different brands and power categories.



Chapter six corresponds to the conclusion part of the thesis where findings of contributory chapters have summarised and comments on different findings have passed. It contains a brief review of the limitations and the generalizations of present contribution; their relative advantages and shortcomings and scope of further work in connection with the present study.



CHAPTER		PAGE
	2.2.2 Probability and Hazard Plots ...	31
	2.3 Discrimination Among the Family of Models	35
	2.4 A Test of Exponentiality Against log-normal	40
	2.5 A Test of weibull Model Against Gamma Model	43
<b>THREE</b>	<b>: METHODS OF STATISTICAL INFERENCES</b>	
	3.1 The Methods of Least Square ...	46
	3.2 The Maximum Likelihood Method ...	49
	3.3 Censoring ...	53
	3.4 Likelihood Under Different Types of Censoring ...	56
	3.5 The Life Table and Actuarial Method ...	65
	3.6 The Product Limit Estimator ...	67
	3.7 The Likelihood Ratio Test ...	73
<b>FOUR</b>	<b>: ON THE ESTIMATION OF LIFE PARAMETERS</b>	
	4.1 Description of the Experiment ...	76
	4.2 Choice of the Model ...	79
	4.3.1 Least Square Estimates ...	88
	4.3.2 Maximum Likelihood Estimates ...	90
	4.4 Goodness of Fit ...	99
	4.5 Comparison of Survival Curve ...	101
<b>FIVE</b>	<b>: STUDY OF COST-BENEFIT FACTORS OF DIFFERENT BRANDS OF INCANDISCENT ELECTRIC BULBS</b>	
	5.1 Introduction ...	105
	5.2 Measurement of Intensity of Light ...	105

CHAPTER		PAGE
5.3	Voltage Intensity Relationship ...	108
5.4	Choice of the Model for Predicting Intensities ...	118
5.5	Comparison of Parameters for Different Brands	119
5.6	Measurement of Power Consumption ...	123
5.7	Voltage-current Relationship ...	124
5.8	Choice of the Model for Predicting Flow of Electricity ...	129
5.9	Comparison of Power Consumption for Different Brands ...	130
5.10	Cost-Benefit Factors ...	131
SIX	: SUMMARISATION OF RESULTS AND CONCLUDING REMARKS	
6.1	Summarisation of Results ...	134
6.2	Concluding Remarks ...	139
APPENDICES	... ..	141
REFERENCES	... ..	159

## CHAPTER ONE

### BRIEF IDEA OF THE PROBLEMS AND THE STATISTICAL TOOLS FOR THE STUDY

#### 1.1 Statement of the Problem

Electric bulb is an important component of our everyday life. We can not think of a modern life in its absence. A few manufacturing industries including one multinational are producing incandiscent electric bulbs of different power in Bangladesh. Volume of production of these industries is small enough to meet the demand of the country. As a result, a large quantity of incandiscent electric bulbs are to import every year to meet the demand of the local market. Taking all together, indigenou and imported, as many as ten brands of incandiscent electric bulbs are available in the local market.

The only specification that an incandiscent electric bulb contains on its body or on the packet is that the maximum power to be consumed in terms of watt/hour for a given voltage of the electricity supplied. The amount of power consumed by an electric bulb is converted into heat and light energy, the proportion of heat and light being dependent on the nature of the raw material used and the technological devices of the manufacturer. The life span and the guarantee period of a bulb are also dependent on these factors. Local manufacturers themselves are lacking of these informations of their products. Multinational and

manufacturers may have these information of their products. Since ours is a sellers market, they do not feel it necessary to provide these informations to the consumers. Rather they avoid, even on inquiry, in the name of business seorecy. The consumers are helpless in choosing the best one but to depend on the attractive colourful advertising propaganda.

## 1.2 Aims and Scope of the Study

The reliability of an electric bulb depends on (a) the average life; (b) the guarantee period i.e., the minimum life span, while the quality depends on (c) the average power consumption or total power consumption during the life span and (d) average intensity of light or the total amount of light it is providing during its life time.

Reliability informations are very much important for large scale consumers in scheduling their procurement or purchase in order to replace the fused bulbs. Advance informations about the expected number of bulbs to fail in a future time interval is essential for the purpose. This is possible if reliability informations are available.

A bulb may have larger guarantee period but smaller average life while other may have smaller guarantee period but larger average life. Similarly a bulb may have lower power consumption. Simultaneously it may provide small amount of light. A bulb may have larger average life while it may provide smaller quantity of light and so on. This implies that neither of these criteria uniquely or partially with some other can

determine the cost effectiveness of an electric bulb in terms of total amount of light it is providing during its life time and the total cost we are to pay for this amount of light unless we consider all the four factors together with the cost of the bulb.

In order to estimate the reliability informations the study is aimed to conduct a life testing experiment on incandiscent bulbs of different manufacturers and of different power in the laboratory for collecting life data under identical conditions. A long time is required for completing such an experiment with a moderate size full sample or with a type-II censored sample. To economize the experiment in terms of both time and money. We have decided to consider a type-I censored sample in this experiment. To complete the study, two other experiments are to be performed for collecting data on the intensity of light and consumption of power. The collected data would be used for estimating the reliability and quality informations. An information-wise comparative study among different manufacturers would be made first followed by prediction of percentage of failures at some future points of time. Finally, a cost benefit analysis would be done to point out the cost effective one in the interest of all the consumers.

Davis (1952) studied the lifetimes of 417 incandiscent electric lamps of 42 different manufacturers, the data being collected from different sources under different conditions. He has assumed and fitted a normal curve to the data but the fit was not good. Epstein and Sobel (1953) used sequential test plan under the assumption of exponential

thing and found a good result. Mandenhall and Hader (1958) used a mixture of exponential distributions in analysing life time of incandescent electric bulbs. Their findings were reasonable. They had proposed a weibull model for further analysis. Investigation to single out a suitable life model to graduate the observed data is a must without which a successful analysis can not be carried on which is also a prime objective of the study.

### 1.3 Concept of Life Testing and Reliability

When we purchase a consumer's goods we expect it to function properly for a reasonable period of time. Unless a consumer is provided with some informations about the average life or guarantee period, he would not, in general purchase a new item, if an old item of known reliability is available. Life testing experiments are designed to measure the average life of the product or to answer such questions as what is the probability that the item will fail in the time interval  $[t, t+h]$ , given that it was working at time  $t$  ?

In a simple life testing experiment a number of items are subjected to tests and the data consists of the recorded lives of all or some of the items. No matter how efficient the manufacturing process is, one or more failures may occur. This failure may be due to:

- i) careless planning, substandard equipment and raw material used, lack of proper quality control, etc.



- ii) random or chance causes. Random failures occur quite unpredictably at random intervals and can not be eliminated by taking steps at the planning, production or inspection stage;
- iii) wear-out or fatigue, caused by the aging of the item. Reliability is a term that is used in our daily life in various contexts which means something or some one we may depend on or count.

In life testing problem we are interested in a quantitative measure of reliability of an item or a system.

The probability of failure free operation of an item in the time interval  $(0, t)$  is known as the reliability of the item upto time  $t$ . This probability is also known as survival probability. In the absence of censoring, the non-parametric estimate of this probability is given by the ratio

$$S(t) \Big|_{t=x} = \frac{\text{No. of items surviving } \geq x}{\text{No. of items initially exposed to test}} \dots (1.1)$$

If  $f(t)$  be the probability density function of the life time  $T$ , of the item, then the probability that the item would fail just before time  $t$  is given by

$$F(t) = \int_0^t f(x) dx \dots \dots (1.2)$$

rate of failure and generally denoted by  $h(t)$  or  $\mu(t)$ . In actuarial and life contingency problems this limiting conditional probability is known as the force of mortality. Thus,

$$\begin{aligned}
 h(t) &= \lim_{x \rightarrow 0} \frac{F(t+x) - F(t)}{x [1 - F(t)]} = \frac{1}{1 - F(t)} \lim_{x \rightarrow 0} \frac{F(t+x) - F(t)}{x} \\
 &= \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad \dots \quad \dots \quad (1.5)
 \end{aligned}$$

$$\Rightarrow f(t) = h(t) S(t) \quad \dots \quad \dots \quad (1.6)$$

(1.5) implies that  $h(t)$  may be a function of  $t$  or may be a constant. If it is a function of time, then it may be an increasing or a decreasing function of time or may be mixed one. Depending on which the distribution may be IFR, DFR or MFR. Again,

$$\begin{aligned}
 h(t) &= \frac{f(t)}{1 - F(t)} = \frac{d F(t)}{dt [1 - F(t)]} \\
 &= - \frac{d}{dt} \log_e [1 - F(t)] \\
 \Rightarrow h(t) dt &= - d \log_e S(t) \\
 \Rightarrow \int_0^t h(u) du &= - \log_e S(t) = H(t) \quad \dots \quad (1.7) \\
 &= \text{cumulative hazard function}
 \end{aligned}$$

$$\Rightarrow S(t) = \exp \left[ - \int_0^t h(u) du \right] \dots \dots \quad (1.8)$$

$$\Rightarrow f(t) = h(t) \exp \left[ - \int_0^t h(u) du \right] \dots \dots \quad (1.9)$$

The mean life of the distribution is given by

$$\begin{aligned} E(t) &= \int_0^{\infty} t f(t) dt = \int_0^{\infty} t h(t) S(t) dt \\ &= S(t) \int_0^{\infty} t h(t) dt - \int_0^{\infty} \left[ \frac{d}{dt} (t, h(t) S(t)) \right] dt \\ &= - \int_0^{\infty} H'(t) \frac{S(t)}{-H'(t)} dt = \int_0^{\infty} S(t) dt \dots \dots \quad (1.10) \end{aligned}$$

Similar way,  $E(T^2) = 2 \int_0^{\infty} t S(t) dt,$

In general,  $E(T^r) = r \int_0^{\infty} t^{r-1} S(t) dt, \dots \dots \quad (1.11)$

Another important parameter of life distributions is the expected residual life i.e., expectation of life after  $t$  period has elapsed and is given by

$$\begin{aligned}
 R(t) &= E(T - t \mid T \geq t) \\
 &= \int_t^{\infty} \frac{(\mu - t) f(u) du}{S(t)} \\
 &= \int_t^{\infty} \frac{S(u) du}{S(t)} \quad \dots \quad \dots (1.12)
 \end{aligned}$$

The conditional probability that the  $i$ th item having been survived upto time  $t$  would fail in the interval  $(t_1, t_2)$  is given by

$$\lambda_{i(t_1, t_2 \mid t_1, t_2 > t)} = \frac{S(t + t_1) - S(t + t_2)}{S(t)} \quad \dots \quad (1.13)$$

The expected number of items to fail in the interval  $(t_1, t_2)$  is

$$N(t_1, t_2 \mid t_1, t_2 > t) = \sum_{i=1}^{n_i} \lambda_i(t_1, t_2) \quad \dots \quad (1.14)$$

Where  $t_1, t_2$  are the minimal and maximum time during which an item can fail with non-zero probability. An item can either fail at time  $t$  or survive. Then the underlying distribution of the probability of failure process for the  $i$ th item is a point binomial with mean  $\lambda_i$  and variance  $\lambda_i(1 - \lambda_i) / n_i$  when  $n_i$  is the No at risk of failure

at time  $t$ . Further, assuming the propensity of failure for items are independent of each other, the pooled variance of the number of failure,  $N$  is given by,

$$\begin{aligned} \text{Var}(N) &= \sum_{i=1}^n n_i \frac{\lambda_i(1 - \lambda_i)}{n_i} \\ &= \sum \lambda_i (1 - \lambda_i) \quad \dots \quad (1.15) \end{aligned}$$

The standard error of the estimates of number of failures in the time interval  $(t_1, t_2)$  is given by

$$\text{S.E.}(N) = \left[ \sum \lambda_i(1 - \lambda_i) \right]^{\frac{1}{2}} \quad \dots \quad (1.16)$$

Thus the stochastic behaviour of the failure process can be studied through either of these four functions (i) the hazard function (ii) the survival or the reliability functions, (iii) the probability density function and (iv) the distribution function.

So far we have considered  $T$  as a continuous random variable. It may happen that the life times are grouped observations or may be some integral numbers, then we are to treat  $T$  as a discrete random variable. Let  $T$  can take on values  $0 \leq t_1 < t_2 < \dots < \dots$  and the corresponding probability function be

$$p(t_j) = P_r(T = t_j), \quad j = 1, 2 \quad \dots \quad \dots \quad (1.17)$$

The corresponding survival function is the

$$S(t) = P_r(T \geq t) = \sum_{J: t_j \geq t} p(t_j) \quad \dots \quad \dots \quad (1.18)$$

The hazard function may be defined as

$$h(t_j) = \Pr(T = t_j / T \geq t_j) = \frac{p(t_j)}{S(t_j)} \quad \dots \quad \dots \quad (1.19)$$

$$\text{But } P(t_j) = S(t_j) - S(t_{j+1})$$

$$\Rightarrow h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)}; \quad j = 1, 2 \quad \dots \quad \dots \quad (1.20)$$

$$\Rightarrow S(t) = \prod_{j: t_j < t} [1 - h(t_j)] \quad \dots \quad \dots \quad (1.21)$$

In a similar way we may write  $H(t) = -\log_e S(t)$  (1.18)

But it is to be noted that

$$H(t) \neq h(t_j) \quad \text{in general}$$

$$j: t_j < t$$

## 1.5 Some Important Life Distributions

Numerous parametric models are used in the analysis of life time data and in problems related to the modeling of aging or failure process. Among univariate models a few particular distributions occupy a central role because of their demonstrated usefulness in a wide range of situations. Foremost in this category are the exponential, Weibull, gamma and log-normal distributions. The motivation for using a particular model in a given situation is often mainly empirical, if having been found that the model satisfactorily describes the distribution of life times in the population under study. This does not of course, imply any absolute "correctness" of the model. Sometimes there are informations about the aging or failure process in a population that suggest a particular distribution, though such informations are rarely specific enough to narrow considerations to just one family of models. This situation will no doubt improve as our understanding of aging and failure processes deepens. Some theoretical motivation for particular models are available in Shooman (1968) and Johnson and Katz (1970).

### 1.5.1 The exponential distribution

The distribution is easy to handle for statistical inference but too much sensitive for prediction. Davis (1952) and Epstein (1958) showed that this distribution can be used as a model to graduate the lifetimes of a wider range of manufactured items. Feigl and Zelen (1965) showed that this distribution can be used in the study of survival or

remission times in chronic diseases. The distribution is characterized by a constant hazard function  $h(t) = \lambda$ ,  $t \geq 0$ ,  $\lambda > 0$ . In other words, the exponential distribution arises in situations where the failure rate appears to be more or less constant. The probability that the item survives for at least time  $t$  is

$$S(t) = \exp - \int_0^t h(x) dx$$

$$= \exp - \int_0^t \lambda dx = e^{-\lambda t} \quad \dots \quad (1.22)$$

$$\Rightarrow f(t) = h(t) S(t) = \lambda e^{-\lambda t} \quad \dots \quad (1.23)$$

There are however, some other elementary considerations which lead to an exponential distribution. These assumptions are stated below:

- 1) The failure of the item in a given interval of time  $[t_0, t_1]$  on the condition that the item works until time  $t_0$  depends only on  $(t_1 - t_0)$ , the length of the time interval and not on  $t_0$ , the position of the time interval.
- 2) On the condition that the item works until time  $t_0$ , the probability that the item will fail in an infinitesimal interval  $[t, t + h]$  is proportional to  $h$  except for higher order.
- 3) The probability of failure at  $t = 0$  i.e., the instant the test started is zero.



Let  $S(t) = p(T \geq t)$   
 = probability that the item survives for  
 at least time  $t$ .

Let  $\lambda$  be the constant of proportionality in assumption (2).  
 In view of the assumptions (1) and (2), we may write using Kolmogorov  
 equation,

$$S(t+h) = S(t) [1 - \lambda h] + o(h)$$

or, 
$$\frac{S(t+h) - S(t)}{h} = -\lambda S(t) + o(h)/h$$

Taking limit  $h \rightarrow 0$  on both sides, we have

$$\frac{d S(t)}{dt} = -\lambda S(t)$$

or, 
$$\frac{d S(t)}{S(t)} = -\lambda dt$$

or, 
$$d \log_e S(t) = -\lambda dt$$

$$\Rightarrow S(t) = A \exp(-\lambda t) \quad \dots \quad \dots \quad (1.24)$$

Using initial condition for assumption (3), we have,  $S(0) = 1 = A$

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

$$\Rightarrow f(t|\lambda) = \lambda e^{-\lambda t} \quad \dots \quad \dots \quad (1.25)$$

The distribution with  $\lambda = 1$  is called the standard exponential distribution. Its probability density function is shown in Figure below.

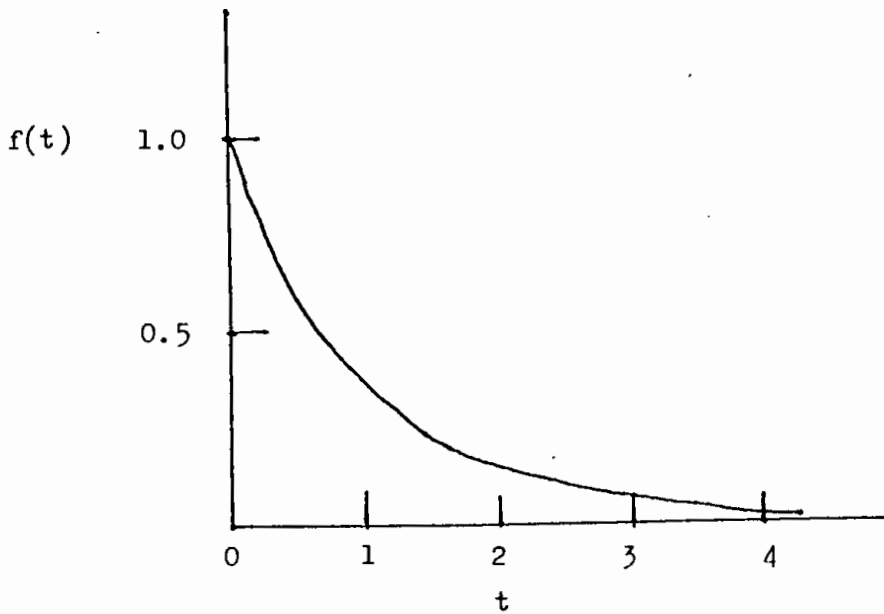


Figure : Standard exponential p.d.f.

### 1.5.2 Some interesting properties of exponential distribution

The mean and variance of the distribution given in (1.20) or (1.22) are  $\frac{1}{\lambda}$  and  $\frac{1}{\lambda^2}$  respectively. The pth quantile is given by

$$P(T \leq t_p) = p$$

$$\text{or, } \int_0^{t_p} \lambda e^{-\lambda t} dt = p$$

$$\text{or } \lambda \left[ -\frac{1}{\lambda} e^{-\lambda t} \right]_0^{t_p} = p$$

$$\text{or, } 1 - e^{-\lambda t_p} = p$$

$$\text{or, } e^{-\lambda t_p} = 1 - p$$

$$\text{or, } -\lambda t_p = \log_e (1 - p)$$

$$\Rightarrow t_p = \frac{-1}{\lambda} \log_e (1 - p) \quad \dots \quad \dots \quad (1.26)$$

The distribution is memoryless, that is, if an item has survived  $t$  hours, then the probability of its surviving an additional  $h$  hours is exactly the same as the probability of surviving  $h$  hours of a new item.

Mathematically,

$$P [ T \geq t + h \mid T \geq t ] = \frac{\int_{t+h}^{\infty} \lambda e^{-\lambda x} dx}{\int_t^{\infty} \lambda e^{-\lambda x} dx}$$

$$= e^{-\lambda h} = P(T \geq h) \quad \dots \quad \dots \quad (1.27)$$

The inter failure times are independent and identically distributed as exponential. Let  $t_1 \leq t_2 \leq \dots \leq t_r$  be the first  $r$  ordered observations of a random sample of size  $n$  for the exponential distribution.

$$\text{Let } w_i = (n - i + 1) (t_i - t_{i-1}), \quad i = 1, 2 \dots r \dots \quad (1.28)$$

with  $t_0 = 0$

The jacobian of transformation is

$$\frac{\partial(w_1, w_2, \dots, w_r)}{\partial(t_1, t_2, \dots, t_r)} = \frac{n!}{(n-r)!} \dots \dots \quad (1.29)$$

$$\begin{aligned} \Rightarrow f(w_1, w_2, \dots, w_r) &= \frac{(n-r)!}{n!} \prod_{i=1}^r f(t_i) \\ &= \frac{(n-r)!}{n!} \frac{n!}{(n-r)!} \lambda^r \left( \prod e^{-\lambda t_i} \right) \left( e^{-\lambda t_r} \right)^{n-r} \\ &= \lambda^r e^{-\lambda [\sum t_i + (n-r)t_r]} \\ &= \lambda^r \exp\left[-\lambda \sum_{i=1}^r w_i\right] \\ &= \prod_{i=1}^r \lambda e^{-\lambda w_i} \dots \dots \quad (1.30) \end{aligned}$$

$\Rightarrow w_i$ 's are independently identically distributed as exponential.

Since  $w_i$  is a linear function of the inter failure time between  $i$ th and  $(i-1)$ th failure so does the inter failure times.

### 1.5.3 The two-parameter exponential distribution

So far we have discussed the negative exponential distribution excluding the threshold parameter or guarantee time, that is, a time say  $\mu \geq 0$  before which it is assumed that an item or a system can not fail. In many situations inclusion of such a parameter becomes essential. It can be done by mere replacing the life time  $t$  by  $t' = t - \mu$ , while  $t'$  satisfies the restriction  $t' \geq 0$ . For example, if we replace  $t$  by  $t' = t - \mu$  in (1.20) or (1.22), we get

$$f(t/\lambda) = \lambda e^{-\lambda(t-\mu)}, \quad t \geq \mu \quad \dots \quad \dots \quad (1.31)$$

The mean and survivor function of (1.31) is  $\mu + \frac{1}{\lambda}$  and  $\exp[-\lambda(t-\mu)]$  respectively which are different from the mean and survivor function (1.23) or (1.25) but variance and other properties remains unaffected by inclusion of the guarantee time or threshold parameter.

### 1.5.4 The Weibull distribution

Historically though exponential is the first widely used lifetime distribution model. But the assumption of a constant hazard function is too much restrictive one. Later realizations that many inferences are sensitive to departures from the exponential model has lead to greater caution in the use of exponential model. Weibull (1951) investigated a model which is, perhaps, the most widely used lifetime distribution now-a-days because of its flexibility in graduating constant,

increasing and decreasing failure rate life time data. The hazard function of the distribution is characterized by

$$h(t) = \alpha \beta (t)^{\beta-1}, \quad \alpha > 0, \beta > 0, t \geq 0 \quad \dots (1.32)$$

This distribution includes exponential as a special case for  $\beta = 1$ , i.e.  $\beta=1$ ,  $h(t)=\alpha$ , a constant. If  $\beta > 1$ ,  $h(t)$  is an increasing function of time  $t$ , while it is a decreasing function for  $\beta < 1$ .

The cumulative hazard function of the distribution in the time interval  $(0, t)$  is given by

$$H(t) = \int_0^t \alpha \beta (\alpha x)^{\beta-1} dx = (\alpha t)^\beta \quad \dots (1.33)$$

$$\Rightarrow S(t) = \exp [-(\alpha t)^\beta] \quad \dots (1.34)$$

$$\Rightarrow f(t) = \alpha \beta (\alpha t)^{\beta-1} \exp [-(\alpha t)^\beta] \quad \dots (1.35)$$

The  $r$ th raw moment of this distribution is,

$$\begin{aligned} \mu_p &= \int_0^\infty t^{p-1} S(t) dt \\ &= \int_0^\infty t^{p-1} \exp [-(\alpha t)^\beta] dt \\ &= \alpha^{-p} \Gamma\left(1 + \frac{p}{\beta}\right) \quad \dots \dots (1.36) \end{aligned}$$

Where  $\Gamma_K = \int_0^{\infty} x^{K-1} e^{-x} dx, K > 0$  (1.37)

The mean of the distribution  $\mu_1 = \alpha^{-1} \left[ 1 + \frac{1}{\beta} \right]$  (1.38)

The variance of the distribution,

$$\begin{aligned} \mu_2 &= \mu_2' - \mu_1'^2 \\ &= \alpha^{-2} \left[ \left( 1 + \frac{2}{\beta} \right) - \left\{ \left( 1 + \frac{1}{\beta} \right) \right\}^2 \right] \end{aligned} \quad (1.39)$$

The pth quantile of the distribution is given by

$$t_p = \frac{1}{\alpha} \left[ -\log_e (1-p)^{1/\beta} \right] \dots \dots \dots (1.40)$$

This distribution arises as an asymptotic extreme value distribution (Gambel, 1958). The shape of the probability density function of this distribution depends on  $\beta$  while  $\alpha$  is a scale parameter. Values of  $\beta$  may vary from application to application but the analysis of  $f(t)$  and  $h(t)$  for different values of  $\beta$  shows that  $\beta$  should lie in the range of  $1 < \beta < 3$  (lawless, 1982).

#### 1.5.5 The log-Normal distribution

The log-normal distribution is also a widely used life model. For example, Nelson and Hahn (1972) used this model for analysing failure times of electrical insulations. Whittimore and Altschuler (1976) used it in the study of times to the appearance of lung cancer

in cigarette smokers. When logarithm of life time  $T$  is distributed normally with mean  $\mu$  and variance  $\sigma^2$  then the life time  $T$  is said to have a log-normal distribution with

$$E(T) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \dots \dots (1.41)$$

$$\text{and } V(T) = \exp(2\mu + \sigma^2) (\exp \sigma^2 - 1) \dots \dots (1.42)$$

The probability density function of  $T$  is given by,

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\log_e t - \mu}{\sigma}\right)^2\right], t > 0 \dots (1.43)$$

The log-normal survivor function is given by

$$S(t) = 1 - \Phi\left(\frac{\log_e t - \mu}{\sigma}\right) \dots (1.44)$$

$$\text{where } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \dots (1.45)$$

The hazard function of the distribution

$$h(t) = \frac{f(t)}{S(t)}$$

is not available in closed form. The hazard function  $h(t)$  has  $h(0) = 0$ , increases to a maximum, then decreases.  $h(t) \rightarrow 0$  as  $t \rightarrow \infty$ . The shape of the distribution depends on  $\sigma$ . For small values of  $\sigma$ , the distribution is more or less asymptotic. As the value of  $\sigma$  increases, the distribution becomes closer and closer to the exponential



distribution. In that case, it is difficult to distinguish one from the other. It is a mixed failure rate distribution.

### 1.5.6 The Gamma distribution

Generally, the one parameter gamma distribution is not in use as a life distribution. The probability density function of the two parameter gamma distribution is given by

$$f(t) = \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma k}, \quad t > 0 \quad (1.46)$$

where  $k > 0$  and  $\lambda > 0$  are index and scale parameters of the distribution. Buckland (1964) showed that this distribution can be used in graduating life data in many situations.

The hazard function and the survival function of the distribution is not available in closed form. For integral values of  $k$ , the survival function may be expressed as

$$\begin{aligned} S(t) &= \int_t^{\infty} f(x) dx \\ &= \int_t^{\infty} \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{\Gamma k} dx \\ &= \frac{\lambda^k}{(k-1)!} \int_t^{\infty} x^{k-1} e^{-\lambda x} dx \\ &= \frac{\lambda^k}{(k-1)!} \left[ x^{k-1} \int_t^{\infty} e^{-\lambda x} dx - \int_t^{\infty} \left\{ \frac{d}{dx} x^{k-1} \int_t^{\infty} e^{-\lambda x} dx \right\} dx \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda^k}{(k-1)!} \left[ x^{k-1} \left\{ \frac{\bar{e}^{-\lambda x}}{-\lambda} \right\}_t^\infty - \int_t^\infty (k-1) x^{k-2} \frac{\bar{e}^{-\lambda x}}{-\lambda} dx \right] \\
&= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} + \frac{\lambda^{k-1}}{(k-2)!} \left[ \int_t^\infty x^{k-2} e^{-\lambda x} dx \right] \\
&= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} + \frac{\lambda^{k-1}}{(k-2)!} \left[ x^{k-2} \int_t^\infty \bar{e}^{-\lambda x} dx - \int_t^\infty \left\{ \frac{d}{dx} x^{k-2} \int_t^\infty \bar{e}^{-\lambda x} dx \right\} dx \right] \\
&= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} + \frac{(\lambda t)^{k-2}}{(k-2)!} e^{-\lambda t} + \int_t^\infty \frac{\lambda^{k-2} x^{k-3} e^{-\lambda x}}{(k-3)!} dx \dots \quad (1.47)
\end{aligned}$$

$$= \sum_{i=0}^{k-1} \left( e^{-\lambda t} \frac{(\lambda t)^i}{i!} \right)$$

For  $k=1$ ,  $s(t) = e^{-\lambda t}$

$$\Rightarrow h(t) = \frac{f(t)}{s(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}$$

$$= \lambda, \quad \text{constant} \quad \dots \quad (1.48)$$

$\Rightarrow$  for  $k=1$ , the two parameter gamma distribution reduces to one parameter negative exponential distribution.

For  $\kappa = 2$ ,  $S(t) = e^{-\lambda t} (1 + \lambda t)$

$$\Rightarrow h(t) = \frac{f(t)}{S(t)} = \frac{\lambda (\lambda t) e^{-\lambda t}}{e^{-\lambda t} (1 + \lambda t)}$$

$$= \frac{\lambda^2 t}{1 + \lambda t} = \frac{\lambda^2}{\lambda + \frac{1}{t}} \dots \dots \quad (1.49a)$$

$\Rightarrow$  As  $t$  increases,  $h(t)$  also increases for  $\kappa = 2$ ;

For  $\kappa = 3$ ,  $h(t) = \frac{\lambda (\lambda t)^2}{1 + \lambda t + (\lambda t)^2} = \frac{\lambda^3}{\lambda^2 + \frac{\lambda}{t} + \frac{1}{t^2}} \dots \dots \quad (1.49b)$

$\Rightarrow$  As  $t$  increases,  $h(t)$  increases for  $\kappa = 3$ . This is true for all values of  $\kappa$ . In a similar way, it can be shown that  $h(t)$  decreases for all values of  $\kappa < 1$ .

Again the limiting value of  $h(t)$  as  $t \rightarrow \infty$  is constant and equal to  $\lambda$  i.e., for  $\kappa = 2$

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \frac{\lambda^2}{\lambda + \frac{1}{t}} = \frac{\lambda^2}{\lambda} = \lambda \quad \dots \dots \quad (1.50a)$$

Similarly for  $\kappa = 3$ ,

$$\begin{aligned} \lim_{t \rightarrow \infty} h(t) &= \lim_{t \rightarrow \infty} \frac{\lambda^3}{\lambda^2 + \frac{\lambda}{t} + \frac{1}{t^2}} \\ &= \frac{\lambda^3}{\lambda^2} = \lambda \quad \dots \dots \quad (1.50b) \end{aligned}$$

In a similar way,

$$\lim_{t \rightarrow 0} h(t) = \lim_{t \rightarrow 0} \frac{\lambda^2}{\lambda + \frac{1}{t}} = \frac{\lambda^2}{\alpha} = 0 \quad \dots \quad (1.51)$$

The m.g.f of the distribution is

$$\begin{aligned} M(\theta) &= \int_0^{\infty} \frac{e^{\theta t} \lambda (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma k} dt \\ &= \left(1 - \frac{\theta}{\lambda}\right)^{-k} \quad \dots \quad (1.52) \end{aligned}$$

$$\Rightarrow \mu_r = k(k+1)(k+2) \dots (k+r-1) \quad \dots \quad (1.53)$$

This distribution does fit a wide variety of life time data but difficult to work with it than those of Weibull model because its survival and hazard functions are not available in closed form. The gamma distribution also may arise for a mixture of exponential distribution. The hazard function of gamma distribution is very close to that of Weibull model except at the extreme right i.e., for very large values of T. So, it is very difficult to distinguish one with the other in case of type - 1 censored data.

## CHAPTER TWO

### ON THE SELECTION OF A LIFE TESTING MODEL

#### 2.1 Problems in Choosing a Life Model

Statistical analysis of life data is a topic of importance in view of its wide application in Engineering, Industry, Commerce, Bio-medical studies, Defence, Manpower planning and other fields. A substantial number of potential life models have been proposed so far but all models are not equally suitable to explain each and every set of data. Some of the models are so close that it is very difficult to distinguish one from the other with a small set of censored data. For example, log-normal distribution with scale parameter greater than two is hardly distinguishable from exponential distribution. Similarly, when coefficient of variation is less than one, Weibull and gamma distribution is difficult to distinguish. Moreover, apart from time trend, failure times may be affected by some other covariates especially in Bio-medical and Manpower studies where patients and personnels are subjected to individual variations. These effects when imposed on the probability density function (p.d.f) causes fluctuations in the p.d.f and creates confusion in making decision.

Above all, most of the life data are heavily censored while tests available for discriminating among the models are sensitive to sample size, the number of failed items and the type of censoring. Again,

choosing a model is also related to availability of appropriate statistical tools. Taking all together, it is a tedious job to choose a suitable model which can explain and interpret an observed life data reasonably.

Among the models available for life testing, the exponential model with constant poisson intensity 'h' is by far the most simple and the easiest to handle statistically. Hence, first step in choosing a life model to explain a set of observed data is to see whether the set of data conforms to an exponential density or not. In general, life data are available in censored form, while tests available for censored data are still not sufficient in all situations. Tests of exponentiality due to Chen, Hollander and Langberg (1983), Hollander and Proschan (1979) are against a fixed alternative. The process of embedding exponential into Weibull or Gamma distribution and to test for the shape parameter to be equal to unity, due to Crow (1974) is too much complex, because no good method is available for estimating the shape parameter of the said distributions for type-I and randomly censored data. Moreover, such tests are not valid for small  $r$ , the number of failed items. Omnibus tests due to Pettit (1977) based on an empirical distribution function may be applied to type-I and type II censored data but these are also limited to large  $n \leq r/2$ , where  $n$  is the total number of items under study. Tests due to Gail and Gaswirth (1978) based on Gini statistics are applicable to type II censored data only. Test proposed by Barlow et al. (1972) and modified by Bain Engelhardt and Wright (1985) cannot

be used for randomly censored data. This test also fails to discriminate between exponential and log-normal for type II censored data. Even for type I censored data, if the censoring time is large enough, this test may give misleading result. This may also happen if the failure times are affected by covariates other than trend. In fact, there is no clear cut test to discriminate between exponential and mixed failure rate models. It is also difficult to discriminate among the members within a family of distributions.

## 2.2 The Plotting Procedures

### 2.2.1 Estimated survivor or Hazard function plotting

Plots of estimated survivor or cumulative hazard function provide useful pictures of univariate lifetime data, as well as information on the underlying life distribution. They can be used for informal checks on the appropriateness of a model and for obtaining parameter estimates within a model. Plots of  $\hat{H}(t) = -\log_e \hat{S}(t)$  against time  $t$  have been already suggested for help in assessing the shape of the hazard function in a distribution. Similar plots can often be used to help assess whether a specific parametric family of models is reasonable. The basic idea is to make plots that should be roughly linear if the proposed family of models is appropriate, since departures from linearity can be readily appreciated by eye. Suppose, for example, that the

possibility of an underlying exponential distribution is being considered. The survival function for the exponential distribution is given by

$$\begin{aligned}
 S(t) &= e^{-\lambda t} \\
 \Rightarrow \log_e S(t) &= -\lambda t \\
 \Rightarrow -\log_e S(t) &= \lambda t \quad \dots \quad \dots \quad (2.1)
 \end{aligned}$$

Therefore, if  $-\log_e \hat{S}(t)$  is plotted against  $t$ , the resultant graph should be roughly linear and passing through the origin, if an exponential model is appropriate. Similarly, the survival function for the exponential distribution having the threshold parameter or the guarantee time  $\mu$  is given by

$$\begin{aligned}
 S(t) &= e^{-\lambda(t - \mu)} \\
 \Rightarrow \log_e S(t) &= -\lambda(t - \mu) \\
 \Rightarrow -\log_e S(t) &= +\lambda t - \lambda\mu \\
 &= \lambda t - \lambda_0 \quad \dots \quad (2.2)
 \end{aligned}$$

It shows that if  $-\log_e \hat{S}(t)$  be plotted against  $t$ , the resulting graph would be roughly linear and having an intercept on  $Y$ -axis in negative direction, if a two parameter exponential model is appropriate.

The survival function for the Weibull distribution is given by



$$\begin{aligned}
 S(t) &= e^{-(\alpha t)^\beta} \\
 \Rightarrow -\log_e S(t) &= (\alpha t)^\beta \\
 \Rightarrow \log [-\log_e S(t)] &= \beta \log \alpha + \beta \log t \\
 &= \gamma + \beta \log t \quad \dots\dots (2.3)
 \end{aligned}$$

This implies that if  $\log [-\log_e \hat{S}(t)]$  be plotted against  $\log t$  the resulting graph would be an appropriate straight line having an intercept on the Y-axis in positive direction if the underlying life times suit a two parameter Weibull distribution.

When the plots are appropriately linear, one can obtain graphical estimates of the parameters by fitting a straight line to the plot and calculating the slope and intercept by the method of least square. The types of procedures described above can be used for models in which some transform of a life time  $T$ , say,  $Y = g(T)$ , has a location-scale parameter distribution and the survivor function is available in closed form. That is,

$$\begin{aligned}
 S(t) &= \Pr (T \geq t) \\
 &= \Pr (X > x) \\
 &= S_1 \frac{(x - \mu)}{\sigma} \quad \dots\dots (2.4)
 \end{aligned}$$

where  $t = g^{-1}(x)$ . Then  $S_1^{-1}[S(t)] = \frac{x - \mu}{\sigma}$  is a linear function of  $x = g(t)$ , and a plot of  $S_1^{-1}[\hat{S}(t)]$  Verses  $g(t)$  should be roughly linear if the family of models being considered is reasonable.

The linear hazard rate distribution with hazard function  $h(t) = \alpha + \beta t$  has cumulative hazard function  $H(t) = \alpha t + \frac{\beta t^2}{2}$ . Thus  $t^{-1} H(t) = \alpha + \frac{\beta t}{2}$  is a linear function of  $t$ , and approximately linear plots of  $t^{-1} H(t)$  versus  $t$  should result if the model is a Rayleigh distribution.

### 2.2.2 Probability and hazard plots

When the data are uncensored or type II censored, it is customary to use probability plots, rather than the plots of estimated survival function or the cumulative hazard function. These are similar to the plots of estimated survival function, except that, instead of entire estimated survivor function, a single point is plotted for each life time.

Probability plots in their most common form are used with location-scale parameter models. Let  $T$  is a random variable with distribution function of the form  $F\left[\frac{t - \mu}{\sigma}\right]$ , where  $\sigma$  is a scale parameter and  $\mu$  is a location parameters ( $\sigma > 0, -\infty < \mu < \infty$ ). Let  $t_{(1)} < t_{(2)} < \dots < t_{(n)}$  be the ordered observation in a random sample of size  $n$  for the distribution of  $T$ . A probability plot is a plot of  $t_{(i)}$  against quantities  $m_i = F^{-1}(a_i)$ , where  $a_i$  is a fixed estimate of  $F\left(\frac{t_{(i)} - \mu}{\sigma}\right)$ . Since  $F^{-1}\left[F\left\{\frac{t_{(i)} - \mu}{\sigma}\right\}\right] = \frac{t_{(i)} - \mu}{\sigma}$ , if the stated model is reasonable the plot of the points  $(t_{(i)}, m_i)$  should be roughly linear. In fact, the points should be fairly near the line  $x = \mu + \sigma m$  and thus estimates of  $\mu$  and  $\sigma$  can be obtained from the plots.

The  $a_i$ 's are sometimes referred to as plotting positions. Several choices for the  $a_i$ 's are used in practice but the two most popular are

$$a_i = \frac{(i - 0.5)}{n} \quad \dots \quad \dots \quad (2.5)$$

$$a_i = \frac{i}{n + 1}$$

An alternative is to plot the points  $(t_{(i)}, \alpha_i)$ , where

$$\alpha_i = \sum_{j=1}^i (n - j + 1)^{-1} \quad \dots \quad (2.6)$$

It is shown in Lawless (1982) that

$$E(t_{(i)}) = \frac{\alpha_i}{\lambda}$$

$$\Rightarrow \alpha_i = \lambda E(t_{(i)}) = E(\lambda t_{(i)}) \quad \dots \quad \dots \quad (2.7)$$

Hence this plot should be approximately linear with slope  $\lambda$ .

To facilitate probability plots, special probability graph papers are available for common distributions. The graph papers have a scale based on values of  $F^{-1}(a)$  but labeled with a scale. This saves the trouble of computing  $F^{-1}(a)$ . We need only to plot the points  $(t_{(i)}, a_i)$ .

Another frequently used plot is the so-called hazard plot due to Nelson (1972). This is essentially the same as a probability plot or a point plot obtained from  $\hat{S}(t)$ , except that instead of being based on

the PL estimate, the plots are based on the empirical hazard function defined as,

$$\hat{H}(t) = \sum_{j: t_j < t} \frac{d_j}{n_j} \quad \dots \quad \dots \quad (2.8)$$

where  $d_j$  is the number of failure at  $t_j$  and  $n_j$  is the risk set just prior to  $t_j$ . The usual procedure is either to plot  $\hat{H}(t)$  or a transform of it in full or to base a point on  $\hat{H}(t)$ . When the data are type II censored, the observed lifetimes  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  are the first  $k$  lifetimes in a sample of size  $n$  and the number of items at risk just prior to  $t_{(j)}$  is  $n_j = n - j + 1$ . This gives,

$$\widetilde{H}(t_{(i)} + 0) = \sum_{j=1}^i (n - j + 1)^{-1}, \quad i = 1, 2, \dots, k \quad \dots \quad (2.9)$$

Plots involving  $\widetilde{H}(t)$  are used in exactly the same way as plots involving  $\hat{S}(t)$ , the connection being obvious in view of the fact that  $H(t) = -\log_e S(t)$ . For one parameter exponential distribution,

$$H(t) = \lambda t \quad \dots \quad \dots \quad (2.10)$$

For an exponential distribution with a threshold parameter or, guarantee time,

$$\begin{aligned}
 H(t) &= \lambda(t - \mu) \\
 &= -\mu\lambda + \lambda t \\
 &= \lambda_0 + \lambda t \quad \dots \quad \dots \quad (2.11)
 \end{aligned}$$

For a two parameter Weibull distribution,

$$\begin{aligned}
 H(t) &= (\alpha t)^\beta \\
 \Rightarrow \log H(t) &= \beta \log \alpha + \beta \log t \\
 &= \beta_0 + \beta \log t \quad \dots \quad \dots \quad (2.12)
 \end{aligned}$$

Hazard plots can be made on ordinary probability paper, but since  $H(t) = -\log_e S(t)$ , values of  $\widetilde{H}(t)$  need to be converted to values of  $\widetilde{S}(t) = \exp [ -\widetilde{H}(t) ]$  to correspond to the a scale on the paper.

To enable hazard plots to be made with as few calculations as possible, commercial hazard plotting papers have been prepared for certain common distributions. The scales on the paper are set up so that the points  $(t, H(t))$  give a straight line, and thus if the assumed model is appropriate, a plot of the points  $[t_{(i)}, \widetilde{H}(t_{(i)} + 0)]$  should be roughly linear.

Generally plots are used informally. It does not make too much difference exactly what plotting positions are chosen and whether plots are based on  $\widehat{S}(t)$  or  $\widetilde{H}(t)$ . Plots based on relatively small number of

observations may give misleading idea about the form of the distribution. An idea of the variability inherent in a plotting procedure is desirable for the investigator, specially about the variances of the quantities plotted. A detailed discussion about plotting is available in Cox (1978).

Plots are not meant as a substitute for formal test and estimation procedures but provide an idea about the model and methods of crude parameter estimates. Thus a complete statistical analysis would require a combination of both informal and formal methods.

### 2.3 Discrimination among the Family of Models

Hazard function is the basic criterion which can adequately discriminate one life model from the other but in practice, it is not measurable. We can only measure cumulative hazards for a specified time interval. For any distribution, the cumulative hazard function  $H(t)$  in the time interval  $(0, t)$  is given by

$$H(t) = \int_0^t h(u) du = - \log S(t) \quad \dots \quad \dots \quad (2.13)$$

where  $S(t)$  is the survivor function in the time interval  $(0, t)$  and  $h(t)$  is the real hazard function but not measurable. We have for exponential distribution

$$h(t) = h_1, \text{ a constant}$$

$$\Rightarrow H(t) = h_1 t \quad \dots \quad \dots \quad \dots \quad (2.14)$$

For Rayleigh distribution, we have

$$h(t) = h_1 + k_1 t$$

$$\Rightarrow H(t) = h_1 t + h_2 t^2 \quad \dots \quad \dots \quad (2.15)$$

$$\text{where } h_2 = k_1/2,$$

For Weibull distribution, we have

$$h(t) = \alpha \beta (\alpha t)^{\beta-1}$$

$$\begin{aligned} \Rightarrow H(t) &= (\alpha t)^\beta \\ &= h_1 t + h_2 t^2 + h_3 t^3 + \dots \end{aligned} \quad (2.16)$$

Similarly for gamma distribution, we have

$$H(t) = \frac{I(k, ht)}{1 - I(k, ht)} = h_1 t + h_2 t^2 + h_3 t^3 + \dots \quad (2.17)$$

and for log-normal distribution,

$$H(t) = \frac{\phi\left(\frac{\log t - m}{6}\right)}{1 - \phi\left(\frac{\log t - m}{6}\right)} = h_1 t + h_2 t^2 + \dots \quad (2.18)$$

$I(k, ht)$  = incomplete gamma function

$\phi\left(\frac{\log t - m}{6}\right)$  = incomplete normal deviate.

For any distribution,  $H(0) = 0 \Rightarrow$  when  $t = 0$ , there is no time hazard. More generally, we can define  $H(t_j - t_i)$  in the time interval  $(t_i, t_j)$ ,  $j > i = 0, 1, \dots, r-1$ ; as

$$\begin{aligned} H(t_j - t_i) &= \int_{t_i}^{t_j} h(t) dt \\ &= \int_0^{t_j} h(t) dt - \int_0^{t_i} h(t) dt \\ &= H(t_j) - H(t_i) \\ &= \log S(t_i) - \log S(t_j) \\ &= \log S(t/t \geq t_j) - \log S(t/t \geq t_i) \quad \dots \quad (2.19) \end{aligned}$$

Thus we can replace  $t$  by the difference  $(t_j - t_i)$  in equation (2.14) - (2.18). Formulation (2.19) has the advantage over (2.13) in the sense that we can have  $\frac{r(r+1)}{2}$  estimates of  $H(t)$  taking  $t_i = 0$  from formulation (2.19), whereas, we can have only  $r$  estimates of  $H(t)$  from formulation (2.13) with respect to  $r$  observed failure times.



This will increase the efficiency of the investigation, especially, when  $r$  is small. We can estimate  $S(t)$  for any observed point in time for any data, irrespective of the nature of censoring and thus for any observed time interval which give us estimates of cumulative hazards in the time interval  $(t_i, t_j)$ , using Kaplan-Meier product limit estimator defined by

$$S(t) = \prod_{J: t_j < t} \frac{n_j - d_j}{n_j} \quad \dots \quad \dots \quad (2.20)$$

where  $n_j$  is the number of individuals at risk at  $t_j$  and  $d_j$  is the corresponding number of failures. Let us denote the estimated cumulative hazard at the  $k$ th failure time in the time interval  $(0, t_j)$  by

$$\hat{H}_{jk} = \log S(t)/t \geq t_j) - \log S(t \leq 0) \quad \dots \quad (2.21)$$

$$j > i = 0, 1, \dots, r-1, k = 1, 2, \dots, r-j+1.$$

Now if we fit a third degree polynomial in  $t_j$  taking  $t_1 = 0$  as

$$\hat{H}_{jk} = h_0 + h_1 t_j + h_2 t_j^2 + h_3 t_j^3 + u_{jk} \quad \dots \quad \dots \quad (2.22)$$

where  $u_{jk}$  is the error term and test the coefficients  $h_1, \dots, h_3$  for zero,

we can reach the following decisions comparing with the theoretical hazard functions:

The observed data conform to

- (i) a simple exponential model or a log-normal model if

$$h_0 = 0, h_1 \neq 0, h_2 = 0 = h_3$$

- (ii) a Rayleigh model or a log-normal model if  $h_0 = 0$ ,

$$h_1 \neq 0 \neq h_2, h_3 = 0$$

- (iii) a Weibull/Extreme value or a Gamma or a log-normal if

$$h_3 \neq 0.$$

- (iv) Since  $H(0) = 0$ , so  $h_0 > 0$  for any case gives us a clue to investigate whether the observed data is affected by covariates other than time hazard. No other method can provide us any information in this regard. This is an advantage of this method.

The method discussed here is being investigated by Mian (1987)

- (v) For two parameter negative exponential model, the survival function is given as,

$$S(t) = \exp - \lambda(t - \mu) \quad , \quad t \geq \mu ; \mu, \lambda > 0$$

$$\Rightarrow H(t) = -\ln S(t) = \lambda t - \lambda \mu$$

$$\Rightarrow H(\mu) = H(t) \text{ at } t = \mu = 0$$

$$\Rightarrow \text{For } t < \mu \quad , \quad H(t) < 0$$

This implies that for a two parameter exponential model, the fitted polynomial is expected to have a negative constant term in addition to equation (2.14). Similar is the case with all other location-scale parameter family of models when the threshold or guarantee time parameter is introduced.

#### 2.4 A Test of Exponentiality Against log-normal

Even with a full sample, it is a difficult problem to distinguish between exponential and log-normal distributions. Often, both the models give equal fit to the data though they differ significantly with respect to hazard function. The log-normal hazards first increases with time and then decreases, whereas, exponential hazard is constant over all possible interval in time and the distribution is known as memoryless distribution.

Tests so far available for discriminating a log-normal distribution from exponential are mostly based on full samples. The test due to Chen, Hollander and Langberg (1983) is applicable only to moderately censored data. In this respect, works of Shapiro and Wilks (1965),

Hegazy and Green (1975), Michael and Schucany (1979) are also worthy of mention. The test available for type II censored data due to Tiku (1981) is also complex one because it requires calculation of variances and covariances of the ordered sample observations.

In discriminating between exponential and log-normal,

Let  $H_0$ : The underlying distribution is exponential.

$H_1$ : The underlying distribution is log-normal.

Under  $H_0$ ,

$$H(t) = ht. \quad \dots \quad \dots \quad (2.23)$$

If  $t_1, t_2, \dots, t_r$  be the 'r' observed failure times, we can estimate 'r' mean values of 'h' corresponding to r observed intervals in time  $(0, t_j)$ ,  $j = 1, \dots, r$  and can test their equality which is equivalent to test the equality of several means. For alternative hypothesis,  $h^s$  are subject to opposite trend so a linear test may give wrong conclusions. As a result, a quadratic test statistic is desired.

Corresponding to equation (2.21), let the estimated average hazards in the time interval  $(0, t_j)$ , at the kth failure time, be denoted by

$\hat{h}_{jk}$  and given by

$$\hat{h}_{jk} = \frac{\hat{H}_{jk}}{t_j - 0} = \frac{\log \hat{S}(t/t \geq t_j) - \log \hat{S}(t/t \geq 0)}{t_j} \quad \dots \quad (2.24)$$

for every value of  $t$ . Thus mean hazard in the interval  $(0, t_j)$  is given by

$$\bar{h}_j = \sum_{k=1}^{r-j+1} \hat{h}_{jk} / (r-j+1). \quad \dots \quad \dots \quad (2.25)$$

Hence mean hazard taking all possible intervals ( $j = 1, 2, \dots, r$ ) is given by

$$h = \frac{1}{r} \sum_{j=1}^r \sum_{k=1}^{r-j+1} \hat{h}_{jk} / (r-j+1) \quad \dots \quad \dots \quad (2.26)$$

which is the estimated constant hazard under  $H_0$ . The cumulative hazard under  $H_0$  in the interval  $(0, t_j)$  is given by

$$\widetilde{H}_j = \bar{h} t_j \quad (j = 1, 2, \dots, r) \quad \dots \quad \dots \quad (2.27)$$

The sum of squared deviations from the estimated mean cumulative hazard  $\widehat{H}$  under  $H_0$  is given by

$$\sum_{j=1}^r (r-j+1) (\widehat{H}_j - \widetilde{H}_j)^2 \quad \dots \quad \dots \quad (2.28)$$

where

$$\widehat{H}_j = \frac{1}{r-j+1} \sum_{k=1}^{r-j+1} \hat{H}_{jk} \quad \dots \quad \dots \quad (2.29)$$

With  $(r - 1)$  degrees of freedom. The sum of squares of errors due to regression fitting is given by

$$\sum_{j=1}^r \sum_{k=1}^{r-j+1} (\hat{H}_{jk} - \hat{H}_j)^2 \quad \dots \quad \dots \quad (2.30)$$

with  $\frac{r(r+1)}{2} - 1$  degrees of freedom.

Now let us define a test statistic as

$$F = \frac{r(r+1) - 2}{2(r-1)} \sum_{j=1}^r \frac{(r-j+1) (\hat{H}_j - \widetilde{H}_j)^2}{\sum_{k=1}^{r-j+1} (\hat{H}_{jk} - \hat{H}_j)^2} \quad \dots \quad (2.31)$$

The statistic  $F$  is distributed as Snedecor's<sup>8</sup> variance ratio  $F$  with  $(r - 1)$  and  $\frac{r(r+1)}{2} - 1$  degrees of freedom. Smaller values of  $F$  favour  $H_0$ , while the higher values favour  $H_1$ . This test is being investigated by Mian et. al (1987).

## 2.5 A Test of Weibull Model against Gamma Model

Works due to Mann and Fertig (1975), Littel et al (1979), Tiku and Singh (1981) for discriminating a Weibull model from a Gamma model are worthy of mention. But all these tests are very clumsy and inferences cannot be drawn without expert knowledge and use of special

tables. We require a very simple test that can be performed with easy and simple to handle.

Let

$H_0$  : The underlying distribution is Weibull

$H_1$  : The underlying distribution is Gamma.

Under  $H_0$ , the survivor function

$$S(t) = \exp - (\lambda t)^\beta \quad \dots \quad \dots \quad (2.32)$$

$$\therefore -\log S(t) = (\lambda t)^\beta$$

$$\text{or } \log [-\log S(t)] = \beta \log \lambda + \beta \log t$$

$$\text{Let, } Y_t = \log [-\log S(t)],$$

$$\beta_0 = \beta \log \lambda, \quad Z_t = \log t$$

Now let us fit the linear model

$$Y_t = \beta_0 + \beta Z_t + E_t \quad (E_t = \text{error term}) \quad \dots \quad (2.33)$$

by ordinary least square method and test the goodness of fit. If the fit is not good, we can directly reject  $H_0$ . If fit is good, then we can generate survival probabilities via estimated hazards for the Weibull model using the estimated parameters from the linear regression fitting and rest the departure of the generated survival probabilities

from the observed survival probabilities for significance. If the departure is statistically significant, we can reject  $H_0$  in favour of  $H_1$ . The sum of squares of derivations due to linear regression is given by

$$\sum (y_t - \hat{\beta}_0 - \hat{\beta} z_t)^2 \quad \dots \quad \dots \quad (2.34)$$

where  $\hat{\beta}_0$  and  $\hat{\beta}$  are the ordinary least square estimators of  $\beta_0$  and  $\beta$ . The sum of squares of deviations due to model under  $H_0$  is given by

$$\sum_t \left[ \log \left\{ -\log S(t) \right\} - \log \left\{ \log \tilde{S}(t) \right\} \right]^2 \quad \dots \quad (2.35)$$

where  $S(t)$  is the observed survival probability from Kaplan Meier (1958) product limit estimator and  $\tilde{S}(t)$  is the survival probability from the model under  $H_0$ . Now let us define a test statistic as

$$F = \frac{\sum_t \left[ \log \left\{ -\log \hat{S}(t) \right\} - \log \left\{ -\log \tilde{S}(t) \right\} \right]^2}{\sum_t (y_t - \hat{\beta}_0 - \beta z_t)^2} \quad \dots \quad (2.36)$$

$F$  is distributed as Snedecor's variance ratio  $F$  with  $(r-2)$ ,  $(r-2)$  d.f.

The calculated values of  $F$  greater than the theoretical values at the specified level of significance will reject the null hypothesis. The test is being investigated by Mian et. al (1987).



## CHAPTER THREE

### METHODS OF STATISTICAL INFERENCES

#### 3.1 The Method of Least Square

The method of least squares gets its name from the minimization of a sum of squares. As a general principle, it states that if we wish to estimate the vector of parameters  $\theta$  in some expression  $p(x, \theta) = 0$ , where the symbol  $x$  represents an observation, we should choose our estimate  $\theta$  so that

$$\sum_{j=1}^n [p(x_j, \hat{\theta})]^2 \quad \dots \quad \dots \quad (3.1)$$

is minimised. As with any other symmetric principle of estimation, the acceptability of the method of least squares depends on the properties of the estimator to which it leads. Unlike the method of maximum likelihood, it has no general optimum properties to recommend it, even asymptotically. However, in an extremely class of situation, it does have the optimum property even in small samples, that it provides unbiased estimators, linear in the observations, which have minimum variance. This situation is usually describes as the linear model, in which observations are distributed with constant variance about mean values which are linear function of the unknown parameters, and in which observations are all uncorrelated in pairs.

Let  $X_1, X_2, \dots, x_k$  be a set of  $k$  variables which influence the variable  $Y$  and the linear relationship of  $Y$  and  $X_i$ 's are as

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i \quad (3.2)$$

$$i = 1, 2, \dots, n$$

In matrix notation, this linear model may be written as

$$Y = X\theta + U \quad \dots \quad (3.3)$$

where  $Y$  is a  $(n \times 1)$  Vector of observations,  $X$  is a  $(n \times \overline{k+1})$  matrix of known coefficient with  $n > \overline{k+1}$ ,  $\theta$  is a  $(\overline{k+1} \times 1)$  Vector of parameters and  $U$  is a  $(n \times 1)$  Vector of error random variables with

$$E(U) = 0 \quad \dots \quad \dots \quad (3.4)$$

and the dispersion matrix

$$V(u) = E(UU') = \sigma^2 I \quad \dots \quad (3.5)$$

where  $I$  is a  $(n \times n)$  identity matrix. The method of least squares requires that we minimize the scalar sum of squares

$$S = U'U = (Y - X\theta)' (Y - X\theta) \quad \dots \quad (3.6)$$

for variation in the components of  $\theta$ . A necessary condition that

S would be minimized is that

$$\frac{\partial S}{\partial \theta} = 0 \quad \dots \quad \dots \quad (3.7)$$

On differentiation, we have

$$2X' (Y - X\theta) = 0 \quad \dots \quad \dots \quad (3.8)$$

which gives for our least square estimator the Vector

$$\hat{\theta} = (X'X)^{-1} X'Y \quad \dots \quad \dots \quad (3.9)$$

If  $|X'X| = 0$ , then we use the generalized inverse, i.e.,

$$\hat{\theta} = (X'X)^+ X'Y \quad \dots \quad \dots \quad (3.10)$$

or simply,  $\hat{\theta} = X^+Y$

$\hat{\theta}$  is unbiased with  $V(\hat{\theta}) = (X'X)^{-1} \sigma^2$ . Details about the optimum properties of least square estimators are available in Kendall and Stuart, vol. 2 (1961) and in many other texts.

If components of  $\theta$  are found to be significantly correlated with each other, we are to improve the estimator of  $\theta$  as suggested by Hoerl, Kennard and Baldwin (1975) as

$$\hat{\theta} = (X'X + mI)^{-1} X'Y \quad \dots \quad \dots \quad (3.11)$$

where  $I$  is a  $(\overline{k+1}) \times (\overline{k+1})$  identity matrix. The quantity  $m$  is defined as

$$m = \frac{(k+1)s^2}{\hat{\theta} \hat{\theta}} \quad \dots \quad \dots \quad (3.12)$$

where  $s^2$  is the estimated error variance and  $\hat{\theta}$  is the ordinary least square estimator of  $\theta$ .

In reference to life testing and reliability, the cumulative hazard function or its transform, in many cases are linear functions of time or the corresponding transform. This relationship in many cases can be estimated by the method of least squares. Because, the method has the advantage that it requires no distribution function of the life times.

### 3.2 The Maximum Likelihood Method

Let  $x_1, x_2, \dots, x_n$  be a random sample observation from a population having the probability density function  $f(x; \underline{\theta})$ . Where  $\underline{\theta}$  is a parameter vector. The joint probability density function of these  $n$  sample observations, regarded as a function of the unknown parameter  $\underline{\theta}$  is called the likelihood function (L.F) of the sample and is written as

$$L(\underline{\theta}/\underline{x}) = f(x_1/\underline{\theta}) f(x_2/\underline{\theta}) \dots f(x_n/\underline{\theta}) \quad \dots \quad (3.13)$$

The likelihood is the value of a density function for a given sample. So far discrete random variable it is a probability. The principle of maximum likelihood directs us to take as our estimator of  $\theta$  that value, say  $\hat{\underline{\theta}}$ , within the permissible range of  $\underline{\theta}$  which makes the likelihood

function as large as possible. That is we choose  $\hat{\theta}$  so that for any admissible value of  $\theta$ ,

$$L(\hat{\theta} / x) \geq L(\theta / x) \quad \dots \quad (3.14)$$

In cases where the range of the distribution  $f(x, \theta)$  is independent of the parameter or if  $f(x, \theta)$  is zero at its terminal points for all  $\theta$  and  $\theta$  may take any real value in an interval, solution of maximum likelihood estimators become easier. Under these regularity conditions, if the likelihood function be differentiated at least twice with respect to  $\theta$ , then the stationary values of the likelihood function within the interval will be given by the solution of the equation

$$\frac{\partial L(\theta/x)}{\partial \theta} = 0 \quad \dots \quad (3.15)$$

A sufficient but not necessary condition that any of these stationary values, say  $\hat{\theta}$  be a local maximum is that

$$\frac{\partial^2 L(\theta/x)}{\partial \theta^2} < 0 \quad \dots \quad (3.16)$$

If there are more than one local maximum available, we are to choose the largest one provided that there is no terminal maximum of the likelihood function at the extreme permissible values of  $\theta$ .

$L(\theta/x)$  and  $\log_e L(\theta/x)$  have their maximum at the same value of  $\theta$ . In practice, it is often simpler to work with the logarithm

of the likelihood function than with the function itself. Under the regularity conditions,

$$\frac{\partial \log L}{\partial \theta} = \frac{\partial L}{\partial \theta} / L \quad \dots \quad (3.17)$$

Since  $L > 0$  and  $\frac{\partial L}{\partial \theta} = 0$ , we therefore can write,

$$\frac{\partial \log L}{\partial \theta} = 0 \quad \dots \quad \dots \quad (3.18)$$

for which

$$\frac{\partial^2 \log L}{\partial \theta^2} < 0 \quad \dots \quad \dots \quad (3.19)$$

If the parameter vector  $\underline{\theta}$  contains  $k$ -parameters  $\theta_1, \theta_2, \dots, \theta_k$

so that

$$\begin{aligned} L(\underline{\theta} | \underline{x}) &= L(\theta_1, \theta_2, \dots, \theta_k | x_1, x_2, \dots, x_n) \\ &= \prod_{i=1}^n f(x_i | \theta_1, \theta_2, \dots, \theta_k) \end{aligned} \quad (3.20)$$

Then the maximum likelihood estimate of the parameter  $\theta_i, i = 1, 2, \dots$

$k$ , is given by

$$\left. \begin{aligned} \frac{\partial L(\underline{\theta} | \underline{x})}{\partial \theta_i} &= 0 ; i = 1, 2, \dots, k \\ \text{or, } \frac{\partial \log L(\underline{\theta} | \underline{x})}{\partial \theta_i} &= 0 \quad i = 1, 2, \dots, k \end{aligned} \right\} \quad (3.21)$$

Let  $\hat{\theta}$  be a consistent estimator of  $\theta$  and  $E(\hat{\theta}) = t$ . Expanding in Taylor's series, we can write,

$$\begin{aligned}
 0 &= \left( \frac{\partial \log L}{\partial \theta} \right)_{\hat{\theta}} \\
 &= \left( \frac{\partial \log L}{\partial \theta} \right)_t + (\hat{\theta} - t) \left( \frac{\partial^2 \log L}{\partial \theta^2} \right)_{\theta^*} \dots \quad (3.22)
 \end{aligned}$$

Where  $\theta^*$  lies between  $\hat{\theta}$  and  $t$ . Since  $t$  and  $\hat{\theta}$  are consistent estimators of  $\theta$ , in large samples  $\theta^*$  will tend in probability to  $\theta_0$ , the true value of  $\theta$ . Further  $\left( \frac{\partial^2 \log L}{\partial \theta^2} \right)_{\theta_0}$  will tend in probability to its expectation i.e.

$$E \left( \frac{\partial^2 \log L}{\partial \theta^2} \right) = - \frac{1}{\text{var}(\hat{\theta})} \dots \quad (3.23)$$

$$\Rightarrow \hat{\theta} = t + \left( \frac{\partial \log L}{\partial \theta} \right)_t \text{var}(\hat{\theta}) \dots \quad (3.24)$$

Maximum likelihood estimators, though widely used in statistical theory are not unique in general. It may or may not be unbiased. But generally, it is consistent and attains the minimum variance bound under regularity conditions. Sometimes, the likelihood equations become too much complicated that iteration methods must be used. Details about this estimators are available in Kendall and Stuart (1961), Mood, Graybill and Boes (1974) and many other texts.

### 3.3 Censoring

Life testing experiments are usually destructive in the sense that the item are destroyed in running the experiment and cannot be used again. Again time to conduct the experiment is also an important factor because of the rapid change in technology. Sometimes it may happen that before completing a life testing experiment on a particular model of a system, the model is replaced by a new improved and cost-effective model. This is in particular true for electronic systems due to rapid progress in technological devices. This leads us to cut short the experimental time as well the number of items to be observed to fail i.e., instead of observing the life times of all the items in a sample, only a part of it. Such a process is known as censoring. Censoring is almost mandatory in life testing experiments with costly sophisticated items.

By censoring, we mean that in a sample of size  $n$ , a known number of observations is missing at either end (in case of single censoring) or at both ends (in case of double censoring). Life time data often come with such a feature that creates special problems in the analysis of the data. Censoring occurs when exact life times are known for only a portion of the individuals under study; the other portion of the life times are known only to exceed certain values. It means that only a lower (or upper) bound on life time is available for the individuals in the sample censoring is common in life testing experiments because of time limits and other restrictions on data collection. In a life testing experiment, for example, it may not be feasible to continue



experimentation until all items under study have failed. If the experiment is terminated before all have failed, then for items still unfailed at the time of termination only a lower bound on life time is available. This is not to say that there is no information available on their life times of unfailed items but only that the information on them is partial. Let in a sample of size  $n$ , the  $r$  ( $r < n$ ) item have failed from starting of experiment to the time period  $t$ . Here the exact life time of  $(n-r)$  items is not known but their initial time period is available. This  $(n-r)$  items is called censored from  $n$  items.

In fact censoring arises in variety of reasons, and we consequently distinguish among several types of censoring processes in the discussion that follows. The basic problem is to determine the sampling distribution and corresponding likelihood function for a given process and then to determine the properties of statistical methods derived from this. Even in relatively simple situations one has to rely heavily on large sample methods. In complicated situation, it may even be difficult to write down a likelihood.

One should not confuse censoring with truncation. Censoring idea has come for the case of sample i.e., the portion of the sample. But the idea of truncation has come in case of population. In the case of censoring we analyse a part of sample values. But in case of truncation, analysis is involved for the part of population values. By the following example we can clear about the difference between the censoring and truncation. Let  $F(x)$  be the probability distribution

function of the continuous life time variable  $X$  and  $f(x) = \frac{d}{dx} F(x)$  (3.25)

be the density function of  $X$  such that  $F(x) = \int_0^x f(x) dx \dots$  (3.26)

and  $\int_0^{\infty} f(x) dx = 1. \dots \dots \dots$  (3.27). Now if we suppose the restriction that  $X \geq m$ , which is called the truncated distribution the required form of  $f(x)$  could be

$$\begin{aligned} f(x \geq m) &= \frac{f(x)}{\int_m^{\infty} f(x) dx} = \frac{f(x)}{1 - \int_0^m f(x) dx} \\ &= f(x) / [1 - F(m)] = f(x) / S(m) \dots \dots \dots (3.28) \end{aligned}$$

The form of distribution when some values are censored from the sample are as follows. Suppose that there are  $n$  individuals under study and that associated with the  $i$ th individual is a life time  $T_i$  and a fixed censoring time  $L_i$ . The  $T_i$ 's are assumed to be independently identically distributed (i.i.d) with p.d.f.  $f(t)$  and survivor function  $S(t)$ . The exact life time  $T_i$  of an individual will be observed only if  $T_i \leq L_i$ . The data from such a setup can be conveniently represented by the  $n$  pairs of random variables  $(t_i, \delta_i)$  where

$$t_i = \min (T_i, L_i) \text{ and } \delta_i = \begin{cases} 1 & \text{if } T_i \leq L_i \\ 0 & \text{if } T_i > L_i \end{cases} \dots \dots (3.29)$$

That is  $\delta_i$  indicates whether the life time  $T_i$  is censored or not, and  $t_i$  is equal to  $T_i$  if it is observed, and to  $L_i$  if it is not. The joint p.d.f. of  $t_i$  and  $\delta_i$  is

$$\rho(t_i, \delta_i) = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \dots \dots \dots (3.30)$$

Here  $t_i$  is mixed random variable with a continuous and a discrete component.

### 3.4 Likelihood Under Different Types of Censoring

There are various types of censoring. A few are discussed here with their corresponding distribution function and likelihood function.

(i) Type I censoring: In this type of censoring (simple or double) observations are censored at the pre-assigned time. That is, experiments are ran over a fixed time period in such a way that an individual's life time will be known exactly only if it is less than some predetermined value. In such situations the data are said to be type I censored. It is also called "Time censored". Suppose in a life test experiment  $n$  items are (may be) placed on test; but before the experiment starting it decide that this experiment will be stopped (is terminate) after time  $L$  has elapsed. Then the life time of the item is known exactly which is failed within time  $L$ , the items had still not failed, these unfailed items exact life time is not known but only known that their life time is exceed time  $L$ . These items are not failed at time  $L$ , which failure time therefore censored. A type I censored sample is one that arises when individuals  $1, 2, \dots, n$  are subjected to limited periods of observation  $L_1, L_2, \dots, L_n$ . So that an individual's life time  $T_i$  is observed only if  $T_i \leq L_i$ . Where all of the  $L_i$ 's (i.e. the termination time) is equal, we sometimes say that the data are simply type I censored, to distinguish this from the general case. It should be noted that in type I censoring the number of exact life times observed is random in contrast to the case of type II censoring, where it is fixed. Let  $i$ th individual's life time

$T_i$  and a fixed censoring time  $L_i$ . The  $T_i$ 's are assumed to be identically independently distributed (i.i.d) with p.d.f  $f(t)$  and survivor fn  $S(t)$ . Here exact lifetime  $T_i$  of an individual will be observed only if  $T_i \leq L_i$ . The data from such a set up can be conveniently represented by the  $n$  pairs of random variables  $(t_i, \delta_i)$  where

$$\begin{aligned} t_i &= \min \text{ value of } (T_i, L_i) \text{ and} \\ \text{value of } \delta_i &= 1 \text{ if } T_i \leq L_i && \dots\dots (3.31) \\ \text{and } \delta_i &= 0 \text{ if } T_i > L_i \end{aligned}$$

That is  $\delta_i$  indicates whether the life time  $T_i$  is censored or not. The value of " $t_i$ " is equal to  $T_i$  if it is observed and to  $L_i$  if it is not. Now we get the joint p. density function of  $t_i$  and  $\delta_i$  is

$$p(t_i, \delta_i) = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \dots\dots (3.32)$$

Here  $t_i$  is a mixed random variable that is it is mixed with continuous and discrete component. In the discrete part. We get,

$$\begin{aligned} \Pr(t_i = L_i) &= \Pr(\delta_i = 0) \\ &= \Pr(T_i > L_i) \\ &= S(L_i) && \dots\dots (3.33) \end{aligned}$$

For values  $t_i < L_i$ , the continuous p.d.f is

$$\begin{aligned} \Pr (t_i / \delta_i = 1) &= \Pr (t_i/t_i < L_i) \\ &= \frac{f(t_i)}{1 - S(L_i)} \quad \dots \quad (3.34) \end{aligned}$$

In the above case the notation  $\Pr(t_i / \delta_i = 1)$  mean the p.d.f of  $t_i$ , when  $t_i < L_i$ . The joint distn. of  $t_i, \delta_i$  thus has components.

$$\begin{aligned} p_p(t_i = L_i, \delta_i = 0) & \because t_i = L_i \Rightarrow \delta_i = 1 \\ &= p(\delta_i = 0) \\ &= S(L_i) \quad \dots \quad (3.35) \end{aligned}$$

Again

$$\begin{aligned} p(t_i, \delta_i = 1) &= p(t_i / \delta_i = 1) \cdot p(\delta_i = 1), \\ & \quad t_i < L_i \\ &= \frac{f(t_i)}{[1 - S(L_i)]} \times [1 - S(L_i)] \\ &= f(t_i) \quad \dots \quad (3.36) \end{aligned}$$

So from the above expressions

$$\Pr (t_i, \delta_i) = f(t_i)^{\delta_i} S(L_i)^{1 - \delta_i} \quad \dots \quad (3.37)$$

If pairs  $(t_i, \delta_i)$  are independent, the likelihood function will be

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \dots \quad (3.38)$$

From the above L.F we can estimate the unknown parameter which is essential of our analysis.

Type II censoring: In this type of censoring certain fixed number of observations are censored at either end or both ends. That is in this case, number of observations are pre-assigned at the time of experiments. The number of censored values is not a random variable a mathematical constant but the time to fail these pre-assigned number of items is a random variable. Let in an experiment  $n$  items are placed on test for life times. A type II censored sample is one for which only the  $r$  smallest observations in a random sample of  $n$  items are observed ( $1 \leq r \leq n$ ). Experiments involving Type II censoring are often used, for example, in life testing, a total of  $n$  items is placed on test, but instead of continuing until all  $n$  items have failed, the test is terminated at the time of the  $r$ th item failure. Let in an experiment the data consist of the  $r$  smallest lifetimes  $T_{(1)} \leq T_{(2)} \leq T_{(3)} \leq \dots \leq T_{(r)}$ , out of a random sample of  $n$  life times  $T_1, T_2, \dots, T_n$  from the life distn. in question. If  $T_1, T_2, \dots, T_n$  are identically independently distributed  $(i, i, d)$  and have

a continuous distn. with p.d.f is  $f(t)$  and survivor function  $S(t)$ , it follows from general results on order statistics that the joint p.d.f of the order statistics  $T_{(1)} T_{(2)} \dots T_{(r)}$  is

$$g(T_{(1)}, T_{(2)} \dots T_{(r)}, \theta) = \frac{n}{(n-r)!} f(t_{(1)}) \dots f(t_{(r)}) [S(t_r)]^{n-r}, \quad t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(r)} \quad \dots \quad (3.39)$$

The above j.p.d.f gives the likelihood function when it is taken as the parameter  $\theta$ .

$$L(\theta / T_{(1)}, T_{(2)} \dots T_{(r)}) = \frac{n}{(n-r)!} f(t_{(1)}) \dots f(T_{(r)}) [S(t_r)]^{n-r} \quad \dots \quad (3.40)$$

From the above L.F we can determine the unknown parameter  $\theta$ .

Progressive Type II censoring: In this censoring, censoring is completed in just two stages: at the time of the  $r_1$  th failure,  $n_1$  of the remaining  $(n-r_1)$  unfailed items, again selected and removed. The experiment then terminates after  $r_2$  items have failed. That is a generalization of type II censoring is the progressive type II censoring.

At this time there are  $(n - r_1 - n_1 - r_2)$  items still unfailed and in this experiment the  $r_1$  observations of the 1st stage is

$$T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(r_1)}$$

and the observation of the 2nd stage is

$$T_{(1)}^* \leq T_{(2)}^* \leq \dots \leq T_{(r_2)}^*$$

Now the sampling distn. of the two stage observation, is

$$\begin{aligned} g(t_{(1)}, t_{(2)}, \dots, t_{(r_1)}, t_{(1)}^*, t_{(2)}^*, \dots, t_{(r_2)}^*) \\ = g_1(t_{(1)}, \dots, t_{(r_1)}) g_2 \left[ \frac{t_{(1)}^* \dots t_{(r_2)}^*}{t_{(1)}, \dots, (r_1)} \right] \dots \end{aligned} \quad (3.41)$$

We know in the case of type II censoring

$$\begin{aligned} g(t_{(1)}, t_{(2)}, \dots, t_{(r_1)}) \\ = \frac{n!}{(n-r_1)!} f(t_{(1)}) f(t_{(2)}) \dots f(t_{(r_1)}) [S(t_{(r_1)})]^{n-r_1} \dots \end{aligned} \quad (3.42)$$

and the 2nd term of equation (3.41) can be obtained by the left truncated distn. with p.d.f and survivor function

$$\begin{aligned} f_1(t) &= \frac{f(t)}{S(t_{(r_1)})}, \quad t > t_{(r_1)} \\ S_1(t) &= \frac{S(t)}{S(t_{(r_1)})}, \quad t > t_{(r_1)} \quad \dots \end{aligned} \quad (3.43)$$

Respectively.



$T_{(1)}^*, T_{(2)}^*, \dots, T_{(r_2)}^*$  are the  $r_2$  smallest observations in a random sample of size  $(n-n_1-r_1)$  from this truncated distn. . By equation (3.42) the 2nd term of equation (1) is therefore

$$\begin{aligned} & g(t_{(1)}^*, \dots, t_{(r_2)}^* | t_{(1)}, \dots, t_{(r_1)}) \\ &= \frac{(n-r_1-n_1)!}{(n-r_1-n_1-r_2)!} f_1(t_{(1)}^*) f_1(t_{(2)}^*) \dots f_1(t_{(r_2)}^*) [S_1(t_{(r_2)}^*)]^{n-r_1-n_1-r_2}, \\ & \quad t_{(1)}^* \leq t_{(2)}^* \leq \dots \leq t_{(r_2)}^*. \\ &= \frac{(n-r_1-n_1)!}{(n-r_1-n_1-r_2)!} \frac{f(t_{(1)}^*) f_1(t_{(2)}^*) \dots f_1(t_{(r_2)}^*)}{[S(t_{(r_1)})]^{n-r_1-n_1}} [S(t_{(r_2)})]^{n-r_1-n_1-r_2} \dots \quad (3.44) \end{aligned}$$

Because

$$[S_1(t_{(r_2)}^*)]^{n-r_1-n_1-r_2} = \frac{[S(t_{(r_2)})]^{n-r_1-n_1-r_2}}{[S(t_{(r_1)})]^{n-r_1-n_1}} \dots \quad (3.45)$$

Now by putting the values of equation (3.41) we get

$$\begin{aligned} & g(t_{(1)}, t_{(2)}, \dots, t_{(r_1)}, t_{(1)}^*, t_{(2)}^*, \dots, t_{(r_2)}^*) \\ &= \frac{n! (n-r_1-n_1)!}{(n-r_1)! (n-r_1-n_1-r_2)!} f(t_{(1)}) \dots f(t_{(r_1)}) [S(t_{(r_1)})]^{n_1} \times \dots \quad (3.46) \\ & \quad f(t_{(1)}^*) f(t_{(2)}^*) \dots f(t_{(r_2)}^*) [S(t_{(r_2)})]^{n-r_1-n_1-r_2} \end{aligned}$$

This joint p.d.f is the likelihood function when it is taken as the parameter  $\underline{\theta}$ .

$$\begin{aligned} & g(\underline{\theta} | t_{(1)}, t_{(2)}, \dots, t_{(r_1)}, t_{(1)}^*, t_{(2)}^*, \dots, t_{(r_2)}^*) \\ &= \frac{n! (n-r_1-n_1)!}{(n-r_1)! (n-r_1-n_1-r_2)!} f(t_{(1)}) \dots f(t_{(r_1)}) [S(t_{(r_1)})]^{n_1} \\ & \quad f(t_{(1)}^*) f(t_{(2)}^*) \dots f(t_{(r_2)}^*) [S(t_{(r_2)})]^{n-r_1-n_1-r_2} \dots \quad (3.47) \end{aligned}$$

Random censoring: In the case of Random Censoring, the censoring times are often effectively random of an life time experiment. Specifically we assume that the censoring time  $L_i$  for the  $i$ th individual is a random variable with survivor and density function  $S(t)$  and  $G(t)$  respectively and further that  $L_1, L_2, \dots, L_n$  are stochastically independent of each other and of the failure times  $T_1, T_2, \dots, T_n$ . Note that the random censorship model includes the special case of type I censoring, where the censoring time of each individual is fixed in advance, as well as the case where items enter the study at random over time and the analysis is carried out at some prespecified time. This latter situation occurs in some medical studies. Let  $(T_i, L_i)$ ,  $i = 1, 2, \dots, n$  be independent and, as in the case of type I censoring,

Define  $t_i = \text{Min}(T_i, L_i)$

and  $\delta_i = 1$  if  $T_i \leq L_i$

or,  $\delta_i = 0$  if  $T_i > L_i$

The data from observation on  $n$  individuals consists of the pair  $(t_i, \delta_i)$

The p.d.f of  $(t_i, \delta_i)$  is easily obtained if  $f(t)$  and  $S(t)$  are the p.d.f's for  $T_i$  and  $L_i$  than, using the same notation as in the likelihood function of the type I censoring, we have

$$\begin{aligned} \Pr(t_i = t, \delta_i = 0) &= P(L_i = t, T_i > L_i) \\ &= g(t) S(t) \end{aligned} \quad \dots \quad (3.48)$$

$$\begin{aligned} \text{and } \Pr(t_i = t, \delta_i = 1) &= P(T_i = t, T_i \leq L_i) \\ &= f(t) G(t) \end{aligned} \quad \dots \quad (3.49)$$

The above two expression can be written together as way

$$\begin{aligned} &Pr(t_i = t, \delta_i) \\ &= [f(t) G(t)]^{\delta_i} [g(t) S(t)]^{1-\delta_i} \dots \quad (3.50) \end{aligned}$$

and sampling distn. of  $(t_i, \delta_i)$  is

$$\begin{aligned} P(t_i, \delta_i) &= \prod_{i=1}^n [f(t_i) G(t_i)]^{\delta_i} [g(t_i) S(t_i)]^{1-\delta_i} \\ &= (\prod G(t_i)^{\delta_i} g(t_i)^{1-\delta_i}) (\prod f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}) \dots \quad (3.51) \end{aligned}$$

If the  $f^n$ .  $G(t)$  and  $S(t)$  do not involve any parameter than which can be neglected from the above function than we get

$$P(t_i, \delta_i) = \prod_{i=1}^n f(t_i)^{\delta_i} [S(t_i)]^{1-\delta_i} \dots \quad (3.52)$$

Then the likelihood fn is

$$L(\underline{\theta} / t_i, \delta_i) = \prod_{i=1}^n f(t_i)^{\delta_i} [S(t_i)]^{1-\delta_i} \dots \quad (3.53)$$

The above L.F is same of the result for L.F of type I censored sampling. Therefore the random censorship model includes the special case of type I censoring, where the censoring time of each individual is fixed in advance.

Details about censoring mechanism are available in Lawless (1982), Kalbfleisch and Prentice (1980) and other texts.

### 3.5 The Life Table and Actuarial Method

Let  $N$  items are kept under observations for a certain period and out of  $N$ ,  $n$  items are observed to fail within the specified period. Then  $\frac{n}{N}$  is the relative frequency of failure in the specified period. The life table is essentially an extension of the relative frequency table to the case of censored data. With the life table, of course, one emphasises estimation of the conditional probability of failure in an interval, given survival at the beginning of the interval and the probability of surviving past the end of an interval.

Let the time axis be divided into  $(k+1)$  intervals  $I_j = (a_{j-1}, a_j)$   $j = 1, 2, \dots, k+1$  with  $a_0 = 0$ ,  $a_k = T$  and  $a_{k+1} = \infty$  where  $T$  is the upper limit on observations. For each item of a random sample of size  $n$  from some population, suppose that one observes either a life time  $t$  or a censoring time  $L$ . The data are however grouped so that it is only known in which interval a particular item has failed or has censored and not the exact life times or censoring times. The data therefore consists of the number of life times, and

censoring times falling into each of the  $(k+1)$  intervals. In the case of the last interval  $I_{k+1}$ , it can be considered that only lifetimes are in the interval, since all items not failing by time  $T$  must fail some time in  $I_{k+1}$ .

Let the distribution of life times from the population under study has survivor function  $S(t)$  and let us define

$$P_j = S(a_j) = \Pr(\text{an item surviving beyond } I_{j-1})$$

$$\begin{aligned} p_j &= \Pr(\text{an item surviving beyond/it survives beyond } I_{j-1}) \\ &= \frac{P_j}{P_{j-1}} \quad \dots \quad \dots \quad (3.54) \end{aligned}$$

$$q_j = 1 - p_j = \Pr(\text{an item fails in } I_j/\text{it survives } I_{j-1})$$

$$\Rightarrow P_j = P_1, P_2; \dots P_j; j = 1, 2, \dots, k+1 \quad \dots \quad (3.55)$$

This implies that probability of surviving past  $I_j$  is given as the product of conditional probabilities of surviving past intervals upto  $I_j$ , given survival to the beginning of each interval and forms the basis for the approach to life table estimation. If there is no censoring,

$$P_j = \frac{N_{j+1}}{N} \quad \dots \quad \dots \quad (3.56)$$

where  $N_j$  is the number of items at the risk of failure at the beginning of the  $j$ th interval  $I_j$  and is given by

$$N_j = N_{j-1} - D_{j-1} - W_{j-1} \quad \dots \quad \dots \quad (3.57)$$

where  $D_j$  = Number of lifetimes in  $I_j = (a_{j-1}, a_j)$

$W_j$  = Number of censoring times in  $I_j$

If there is no censoring,

$$q_j = \frac{D_j}{N_j - W_{j/2}} \quad \dots \quad \dots \quad (3.58)$$

$$p_j = 1 - q_j$$

The variance of  $P_j$  for censored sample is given as

$$\text{Var}(P_j) = P_j^2 \sum_{i=1}^j \frac{q_i}{P_i (N_i - W_{i/2})} \quad \dots \quad (3.59)$$

If there is no censoring

$$\text{Var} (P_j) = \frac{P_j (1 - P_j)}{N}$$

Details about this method is available in Elandt-Johnson (1977) and Lawless (1982).

### 3.6 The Product Limit Estimator

Graphical representation of the empirical survivor function or the empirical distribution function is an essential way of portraying ungrouped univariate survival data. Non-parametric estimate of the survivor function of the distribution for the life distribution under study is also provided by such graphical representation. If there is

no censored observations in a sample of size  $n$ , the empirical survivor function (E.S.F) is defined as

$$S(t) = \frac{\text{Number of lifetimes observed upto time } t}{n}, \quad t \geq 0 \quad \dots (3.60)$$

This is a step function that decreases by  $\frac{1}{n}$  just after each observed life time if all observations are distinct. More generally, if there are  $d$  lifetimes equal to  $t$ , then ESF drops by  $\frac{d}{n}$  just part  $t$ . When dealing with censored data, some modification is necessary, since the number of life times greater than or equal to  $t$  will not generally be known exactly. The modification due to Kaplan-Meier (1958) is known as product Limit Estimator (PLE) or Kaplan-Meier Estimator. The estimator is defined as follows. Suppose that there are  $n$  items in the sample  $t_1 < t_2 < \dots < t_k$  are  $k$  distinct lifetimes and  $m_j$  items are censored in between  $(t_{j-1}, t_j)$ ,  $j = 1, 2, \dots, k$ . The possibility of more than one failure at  $t_j$  is allowed and let  $d_j$  be the number of items failing at  $t_j$ ,  $J = 1, 2, \dots, k$ . In addition to the observed life times  $t_1, t_2, \dots, t_k$ , there are also censored times  $t_{j1}$  for individuals whose lifetimes are not observed but greater than  $t_j$ . The product limit estimator of the survivor function  $S(t)$  is defined as

$$S(t) = \prod_{j: t_j \leq t} \frac{n_j - d_j}{n_j} \quad \dots \quad \dots \quad (3.61)$$

where  $n_j$  is the number of items at the risk of failure at  $t_j$  i.e., the number of items retaining in the system and uncensored just prior to  $t_j$ ,  $j=1, 2, \dots, k$ . If a censoring time  $t_{j1}$  and a life time  $t_j$  are recorded as equal, the adopted connection is that censoring time is adjusted an infinitesimal amount to the right, so that  $t_{jL}$  is considered to be infinitesimally larger than  $t_j$ . In other words, an item with censoring time equal to  $t_j$  is included in the risk set at  $t_j$ . If the largest observation, the product limit Estimator is defined only upto the last observation. The motivation for PLE is essentially the same as for the survival function in the lifetable method. That is, the PLE is build up as a product and each term in the product can be thought of as an estimator of the conditional probability of surviving past  $t_j$ , given survival till just prior to  $t_j$ ,  $j=1, 2, \dots, k$ . The product limit Estimator is, in fact, a limiting case of the standard life table method, when the number of intervals becomes infinite and length of all intervals except the last one approaches zero. The PLE is always a step function with  $S(0) = 1$ , drops by a factor  $\frac{n_j - d_j}{n_j}$ , immediately at censoring times  $t_{j1}$ . The effect of censoring is, however, felt in the values of  $n_j$  and hence in the sizes of the steps in  $S(t)$ .

Since the survivor function  $S(t)$  is a non-increasing left continuous function, the probability that an item would fail at  $t_j$  is thus given by



$$S(t_j) - S(t_j + 0) \quad \dots \quad (3.62)$$

where  $S(t_j + 0) = \lim_{x \rightarrow 0^+} S(t_j + x)$ ;  $J = 1, 2, \dots, k$

Contribution to the likelihood of a survival time censored at  $t_{j\mathbf{l}}$  is given by

$$P(T > t_{j\mathbf{l}}) = S(t_{j\mathbf{l}} + 0) \quad \dots \quad (3.63)$$

Thus, we obtain

$$L = \prod_{j=0}^k [S(t_j) - S(t_j + 0)]^{d_j} \prod_{\mathbf{l}=1}^{m_j} S(t_{j\mathbf{l}} + 0) \quad (3.64)$$

which given the data, can be viewed as a likelihood function on the space of survivor function  $S(t)$ . The maximum likelihood estimate of  $S(t)$  is  $\hat{S}(t)$  that maximises  $L$ . Clearly  $\hat{S}(t)$  is discontinuous at the observed failure times since otherwise  $L = 0$ . Further, subject to  $t_{j\mathbf{l}} > t_j$ ,  $S(t_j + 0)$  is maximised with the adopted connection

$$S(t_{j\mathbf{l}} + 0) = S(t_j + 0), \quad j = 1, 2, \dots, k; \quad \mathbf{l} = 1, 2, \dots, m_j$$

and  $S(t_{0\mathbf{l}}) = 1$ ,  $\mathbf{l} = 1, 2, \dots, m_0$ . The function  $\hat{S}(t)$  is then a discrete survivor function with hazard components  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$

at  $t_1, t_2, \dots, t_k$  respectively. Thus

$$\hat{S}(t_j) = \prod_{l=1}^{j-1} (1 - \lambda_l) \quad \dots \quad \dots \quad \dots \quad (3.65)$$

$$\hat{S}(t_{j+0}) = \prod_{l=1}^j (1 - \lambda_l) \quad \dots \quad \dots \quad \dots$$

$$\begin{aligned} \Rightarrow L &= \prod_{j=1}^k \left\{ \left[ \prod_{l=1}^{j-1} (1 - \lambda_l)^{d_j} \right] \lambda_j^{d_j} \left[ \prod_{l=1}^j (1 - \lambda_l)^{n_j} \right] \right\} \\ &= \prod_{j=1}^k \lambda_j^{d_j} (1 - \lambda_j)^{n_j - d_j} \quad \dots \quad \dots \quad \dots \quad (3.66) \end{aligned}$$

$$\log L = \sum_{j=1}^k \left\{ d_j \log \lambda_j + (n_j - d_j) \log (1 - \lambda_j) \right\}$$

$$\therefore \frac{\partial \log L}{\partial \lambda_j} = \frac{d_j}{\lambda_j} - \frac{n_j - d_j}{1 - \lambda_j}$$

$$\Rightarrow \hat{\lambda}_j = \frac{d_j}{n_j} \quad \dots \quad \dots \quad (3.67)$$

$$\Rightarrow \hat{S}(t_j) = \prod_{l=1}^j \left( 1 - \frac{d_l}{n_l} \right)$$

$$\Rightarrow \hat{S}(t) = \prod_{j: t_j \leq t} \left( \frac{n_j - d_j}{n_j} \right) \quad \dots \quad \dots \quad (3.68)$$

This shows that PLE is a non-parametric maximum likelihood estimator of the survivor function.

Again

$$\begin{aligned}\hat{S}(t) &= \prod_{j: t_j \leq t} \left( \frac{n_j - d_j}{n_j} \right) \\ &= \prod_{j: t_j \leq t} s_j = \hat{s}_1 \hat{s}_2 \cdots \hat{s}_j \quad \dots\end{aligned} \quad (3.69)$$

where  $s_j$  is the conditional probability of survival at  $t_j$  given the risk set  $n_j$ . Now

$$\begin{aligned}E[\hat{S}(t)] &= E[\hat{s}_1 \hat{s}_2 \cdots \hat{s}_j] \\ &\quad j: t_j \leq t \\ &= E(\hat{s}_1/n_1) E(\hat{s}_2/n_2) \cdots E(\hat{s}_j/n_j) \\ &= s_1 s_2 \cdots s_j \quad \dots\end{aligned} \quad (3.70)$$

This shows that the PLE is an unbiased estimator of the survivor function conditional on the risk set, further,

$$\begin{aligned}V(\hat{S}(t)) &\simeq \sum_{j: t_j \leq t} \left( \frac{\partial \hat{S}(t)}{\partial \hat{s}_j} \right)^2 \text{var}(\hat{s}_j/n_j) \\ &\quad + \sum_{j < k} \frac{\partial \hat{S}(t)}{\partial \hat{s}_j} \frac{\partial \hat{S}(t)}{\partial \hat{s}_k} \text{cov}(\hat{s}_j, \hat{s}_k / \{n_j\}) \\ &= \sum \left( \frac{\hat{S}(t)}{\hat{s}_j} \right)^2 \frac{\hat{s}_j (1 - \hat{s}_j)}{n_j} + 0 \\ &= \hat{S}(t)^2 \sum \frac{(1 - \hat{s}_j)}{n_j \hat{s}_j} \quad \dots\end{aligned} \quad (3.71)$$

$$\begin{aligned}
 &= \hat{S}(t)^2 \sum_j \frac{d_j/n_j}{n_j(1-d_j/n_j)} \\
 &= \hat{S}(t)^2 \sum_j \left( \frac{d_j}{n_j(n_j-d_j)} \right)
 \end{aligned}$$

$$\Rightarrow \lim_{n_j \rightarrow \infty} \text{Var} [\hat{S}(t)] = 0 \dots \dots \dots \quad (3.72)$$

$j: t_j \leq t$

This shows that the PLE is a consistent estimator of the survivor function.

### 3.7 The Likelihood Ratio Test

Let  $x_1, x_2, \dots, x_n$  be a random sample from a distribution with p.d.f  $f(x, \theta)$  where  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  is a vector of unknown parameters taking on values in the parameter space  $\Omega$ . The likelihood function for  $\theta$  is defined as

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) \dots \dots \dots \quad (3.73)$$

The maximum likelihood estimator of  $\theta$  is given by

$$U_i(\theta) = \frac{\partial \log L(\theta)}{\partial \theta_i}, \quad i = 1, 2, \dots, k. \quad (3.74)$$

$U_i(\theta)$ 's are called scores and the  $k \times 1$  vector  $U(\theta) = [U_1(\theta), U_2(\theta), \dots, U_k(\theta)]'$  is called the score vector. It is shown in Cramer (1946) that under mild regularity conditions, the score vector is asymptotically normally distributed with mean 0 and covariance matrix  $I(\theta)$  with entries

$$I_{ij}(\theta) = E \left( - \frac{\partial^2 \log L(\theta)}{\partial \theta_i \partial \theta_j} \right), \quad i, j = 1, 2, \dots, k \quad (3.75)$$

This implies that  $BU(\theta)$  is asymptotically standard  $k$ -variate normal where  $BB' = I(\theta)$ . This means that under the null hypothesis

$$H_0: \theta = \theta_0$$

$$U'(\theta_0) I(\theta_0)^{-1} U(\theta_0) \dots \quad (3.76)$$

is asymptotically  $\chi^2_{(k)}$ . This further implies that under

$$H_0: \theta = \theta_0,$$

$$\Lambda = -2 \log \left[ \frac{L(\theta_0)}{L(\hat{\theta})} \right] \dots \quad (3.77)$$

is asymptotically  $\chi^2_{(k)}$ . If  $\theta$  is partitioned such that

$$\theta = (\theta_1, \theta_2)'$$
 and we consider

$$H_0: \theta_1 = \theta_{10}, \text{ then}$$

$$\Lambda = -2 \log \left\{ \frac{L[\theta_{10}, \tilde{\theta}_2(\theta_{10})]}{L(\hat{\theta}_1, \hat{\theta}_2)} \right\} \quad (3.78)$$

is asymptotically distributed as  $\chi^2_{(p)}$  where  $\theta_1$  is a  $p \times 1$  vector and  $\tilde{\theta}_2(\theta_{10})$  is the maximum likelihood estimate of  $\theta_2$  under  $H_0$ . For survival distributions with censoring, specific assumptions about censoring is required for this test. One approach is to assume that censoring times are distributed across intervals and modify Chi-square in the spirit of standard life table method of estimation or PLE method. This is not particularly attractive. Rather we assume that censoring occurs only at the end of intervals  $I_1, I_2, \dots, I_r$ . Let  $S_j$  be the probability of survival beyond  $I_j$ . To test the best fit, let us consider the hypothesis

$$H_0: S_j = S_{j0}, \quad j = 1, 2, \dots, r$$

$$H_1: S_j \text{'s satisfy only } 0 \leq S_j \leq 1$$

This leads to the likelihood ratio statistic as

$$\chi^2 = -2 \log_e \left[ \frac{L(S_{10}, S_{20}, \dots, S_{r0})}{L(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_r)} \right] \quad (3.79)$$

which follows approximately chi-square distribution with  $k$  d.f. Since PL estimates are non-parametric MLE of the survival function. We can use PL estimate of  $S(t)$  in place of

$$\hat{S}_1, \hat{S}_2, \dots, \hat{S}_r.$$

## CHAPTER FOUR

### ON THE ESTIMATION OF LIFE PARAMETERS

#### 4.1 Description of the Experiment

In order to obtain informations on life parameters, generally manufactured items such as mechanical or electrical components or systems are subjected to life test experiments which involve putting items in question into operation. We did such an experiment in a laboratory in the department of statistics, University of Rajshahi. For our experimental purpose, the electrical net-work was designed and installed by the engineering department, University of Rajshahi under the supervision of a qualified electrical engineer in collaboration with my supervisor. The electrical net work had three parallel lines. Each of the three lines was installed separately from the basper (main switch) and separated by two cutouts. The line were connected by loose wire and set on the table of button holder. The arrangement of electrical network is shown in Figure I. We have pointed out in the statement of the problem in Section 1.1 that as many as ten different brands of incandiscent electric bulbs are available in our local market. Again each brand of bulb has different power, ranging from 5-250 watt in general. It would be better if we could include all brands and all power categories of each brand in our study. But resource constraints compelled us to limit our experiment and consequently our study. We had to include only three brands of bulbs, each with two power categories. For selecting the brands, we have done a preliminary market survey regarding the sell volume of different brands

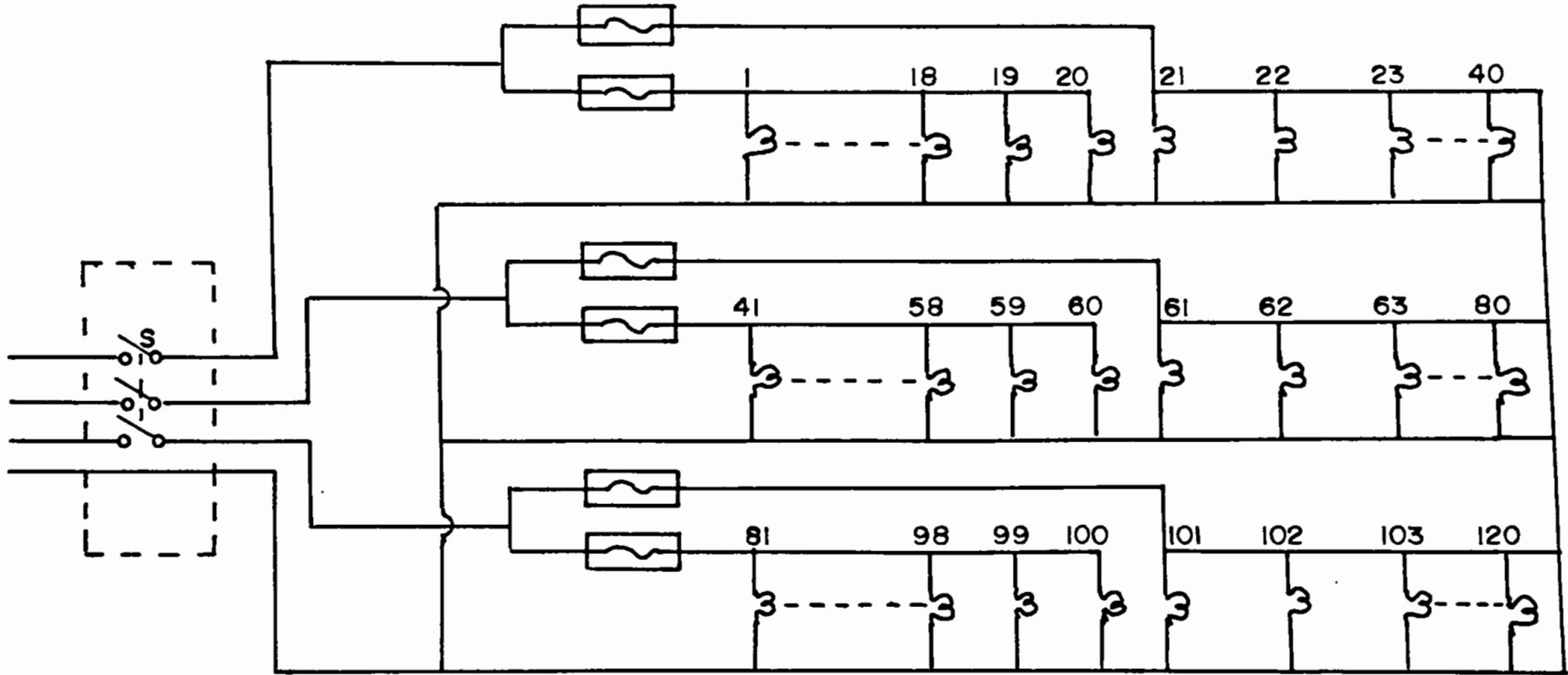


Fig.1. Laboratory setting for the experiment



of incandiscent electric bulbs by power category in Rajshahi. The result of the survey is shown in Appendix I. As per this survey, Phillips has the highest sell volume, Belco is leading in the medium sell volume group while Bengal belongs to the lowest sell volume group. Of the power categories, 60 watt has the highest sell volumes followed by 100 watt. Accordingly, we have included Phillips, Belco and Bengal each with 60 W and 100 W power categories in our experiment as well in the study.

It is mentioned earlier that life testing experiments are destructive one and we were in resource constraint, we put 20 bulbs of each category, in total 120 bulbs, in the experiment. Twenty holder points was set with each cut-out. The holder points were numbered serially 1-120. The bulbs were set in holder points by random allocation. For the purpose, bulbs were placed in identical packets, mixed together and numbered. Two sets of random numbers were drawn simultaneously -- one for holder position and the other for bulbs. Let  $x$  be the random number representing holder positions and  $Y$  be the random number representing bulbs. Then the random set  $(x, y)$  represent the  $y$ th bulb placed in the  $x$ th holder.

Since, life testing experiments are time censoring, we had decided in advance to run the experiment for three calendar months which would produce a type-I censored data. Later on, we encountered a problem. On the eighth day of the experiment, the main switch connecting our experimental net-work was burnt. It took 42 hours to replace the main switch by a new one. In addition, there were 66 recorded hours of electricity failure during the period of experiment. Considering the

electricity failure as a loss time, in total we have recorded 108 hours of loss time. As a result, we have extended our experimental time for 5 days more. It is to be noted that we could not record electricity failure, if there be any, in between 11 p.m. and 7 p.m. Of course, possibility of electricity failure during this period is very little. However, as it is a comparative study, we assume that the effect due to this unobserved loss time, if there be any, would be distributed to all brands and all power categories equally.

We have set our experiment on 29th March, 1989 at 11 p.m. and the experiment was terminated on 4th July, 1989 at 3 p.m. We had no facility to stay in the laboratory for 24 hours observations. We have recorded observed failure times at eight hours interval -- 11 p.m., 7 a.m. and 3 p.m. Failure times were considered at the centre of the interval at which failure occurred. The loss times are shown in Appendix II. The observed failure times and the actual life times are shown in Appendix III for each brand and for each power category separately.

#### 4.2 Choice of the Model

Choice of a model to graduate and analyse a particular set of life data for the purpose of inference is very difficult, since inferences are very much sensitive to underlying models, especially, with censored data. We have discussed the problem in detail in Section 2.1. In the process of selecting a model, the first step is to estimate the survival probabilities by non-parametric methods like the life table method or the

Kaplan-Meier method (PLE). Analysis of data can be done on the basis of non-parametric methods of estimation but it is not very much useful for inference purposes, particularly with censored data. Graphical representation of estimated survival function or its transform may give us some insight about the form of the distribution which we have discussed at length in Section 2.2. We have estimated the survival probabilities at failure points in time for all the six categories of incandescent electric bulbs under our study by the PLE defined in Section 3.6, equation (3.61). Survival probabilities estimated thus are shown in Appendix IV. A graph of  $H(t) = -\ln S(t)$  against time  $t$  is found to be approximately linear having a negative intercept on the Y-axis for all the six categories of life times under investigation. The graphs are shown in figures (2-7). The length of the X-axis between the origin and the point through which the line is passing are varying for the six different categories. Also the angles these lines are making with the positive direction of the X-axis are differing with each other. It is evident from our discussion in Section 2.2 that the underlying distributions belong to some location-scale family of distributions. It also implies that the distributions in question, more or less, belong to the same family differing in their location-scale parameters or to close allies.

To identify the family of distributions to which each of the six categories of lifetimes belongs, we have followed the method of polynomial fitting discussed in Section 2.3. We have tried a third degree polynomial for each category of lifetimes. We found second and third degree coefficients nearly zero and statistically insignificant. A negative

Fig.2: Graph of  $-\ln S(t)$ , against failure time  $t$  for Phillips 60 Watt Bulbs

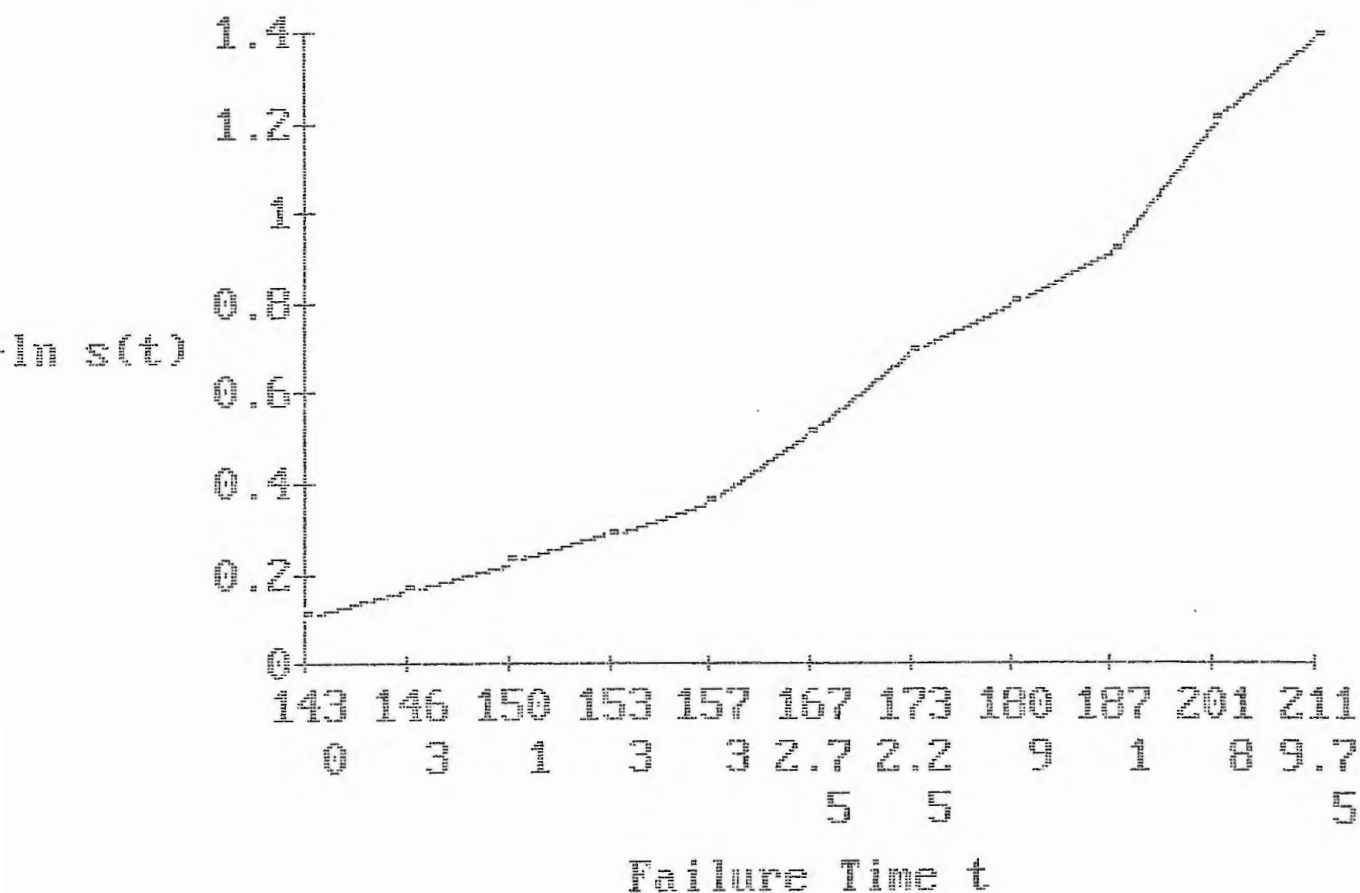


Fig. 3: Graph of  $-\ln S(t)$ ,  
against Failure Time  $t$  for Delco  
60 Watt Bulbs

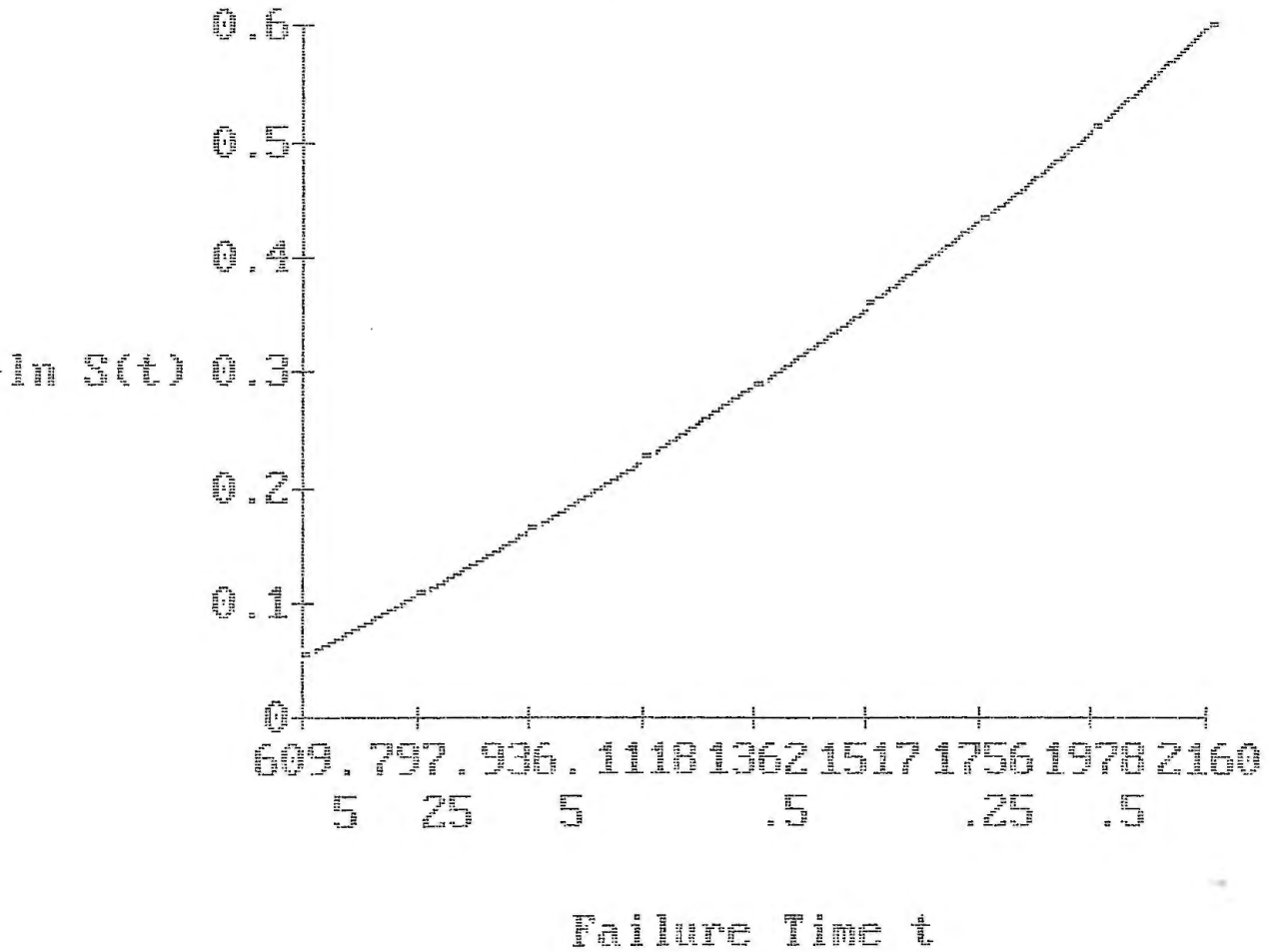


Fig.4: Graph of  $-\ln S(t)$ , against failure time  $t$  for Bengal 60 Watt Bulbs

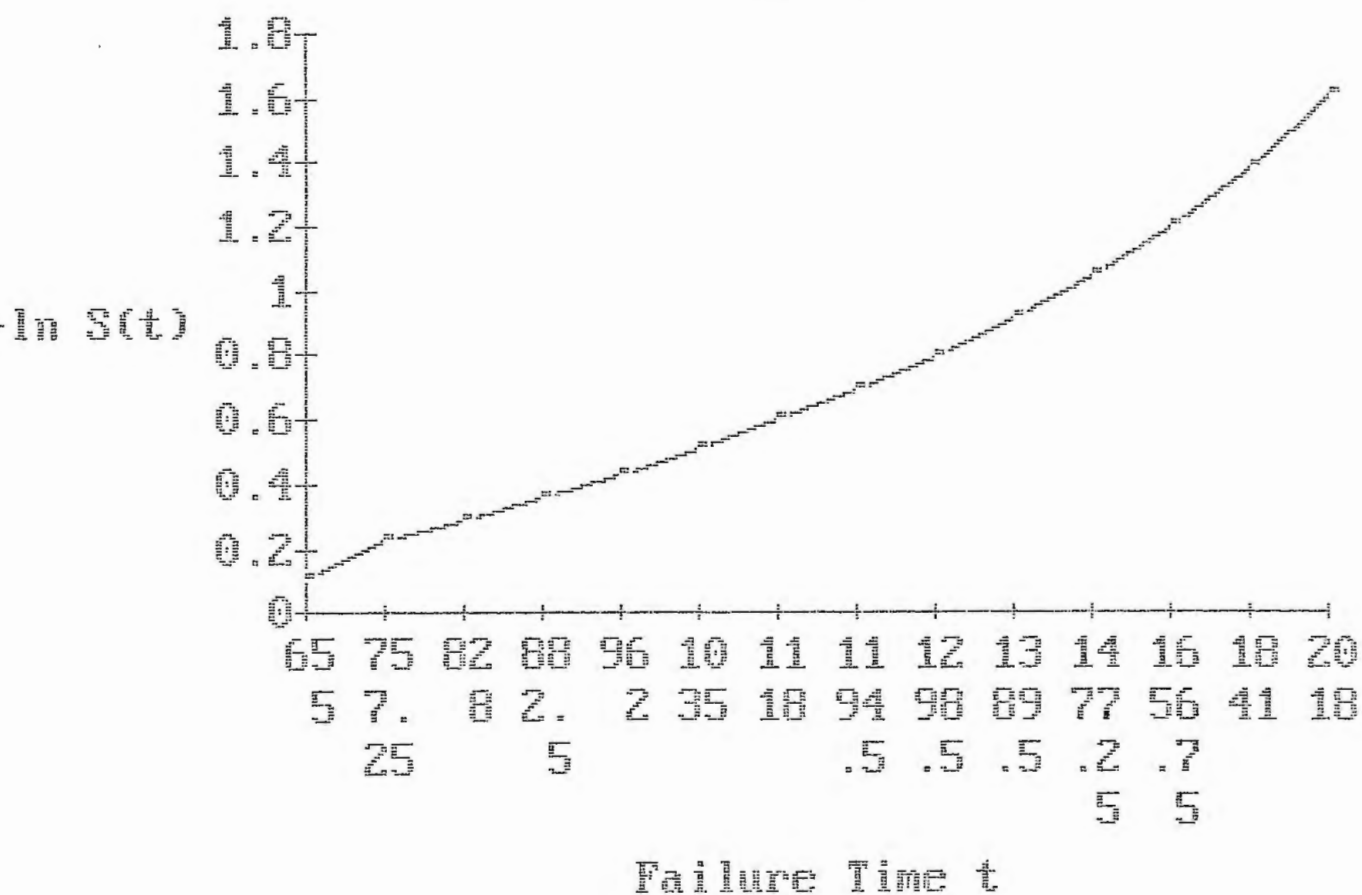


Fig.5: Graph of  $-\ln S(t)$ ,  
against Failure time  $t$  for  
Phillips 100 Watt Bulbs

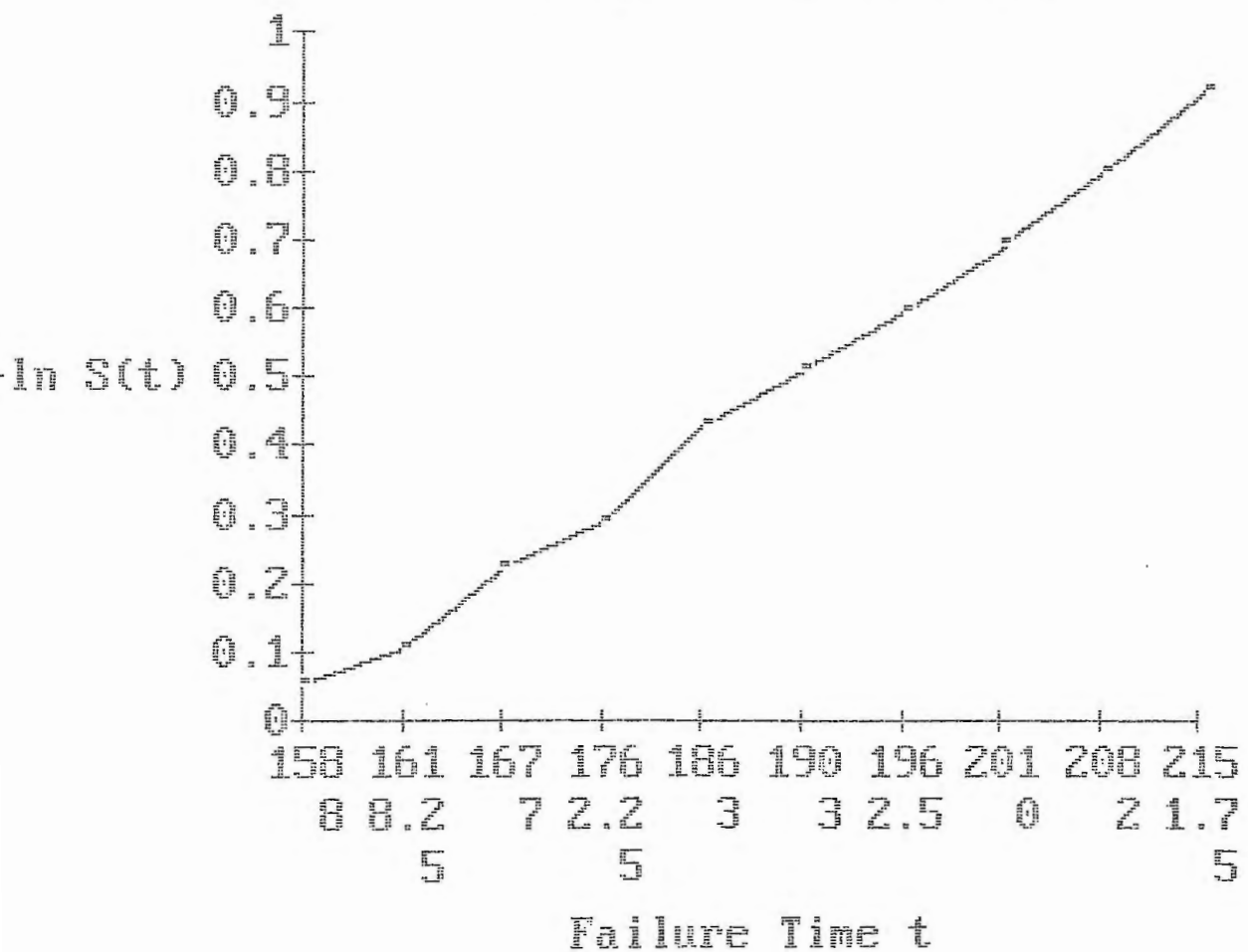


Fig. 6: Graph of  $-\ln S(t)$ ,  
 against Failure Time  $t$  for Seico  
 100 Watt Bulbs

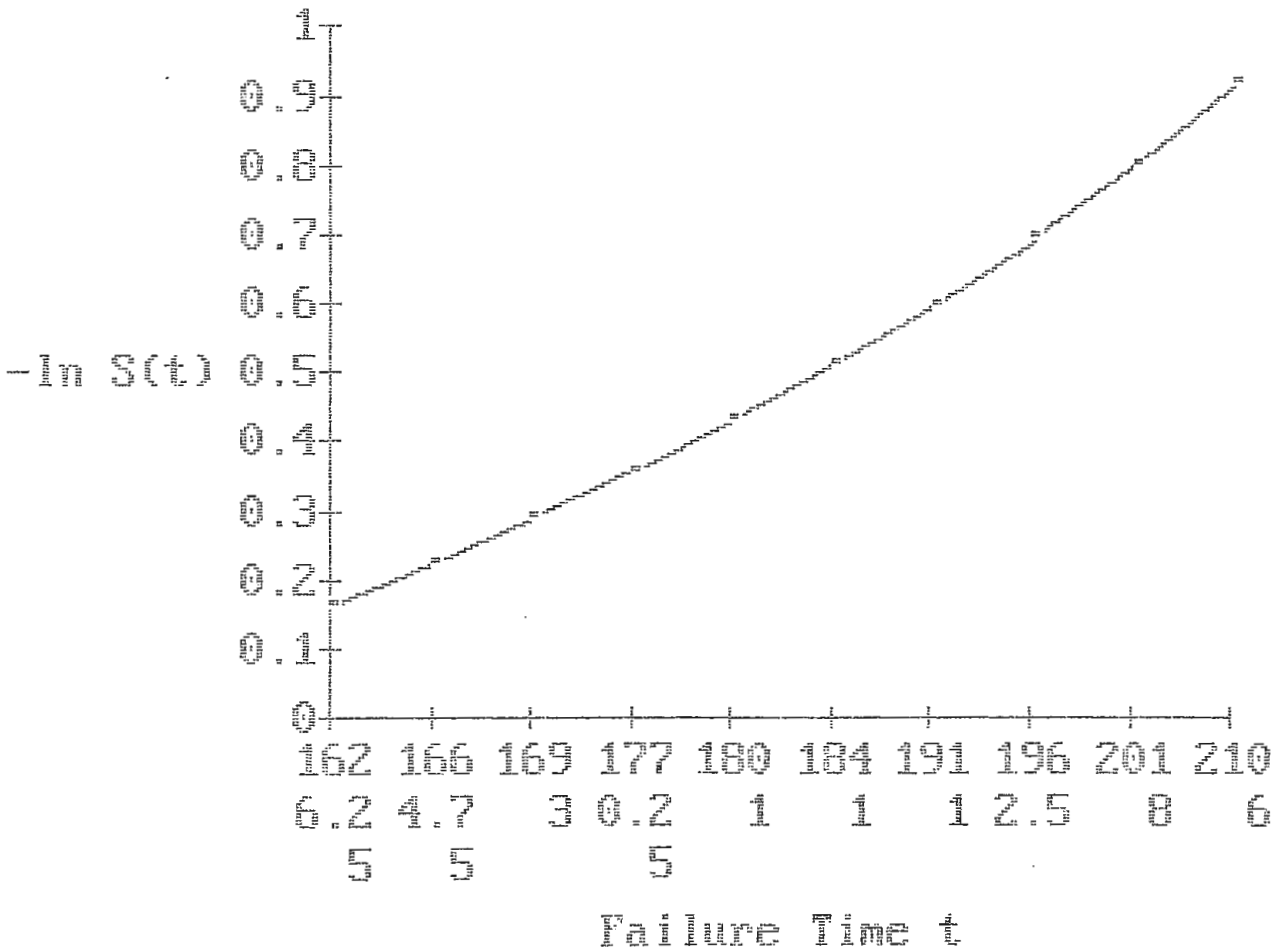
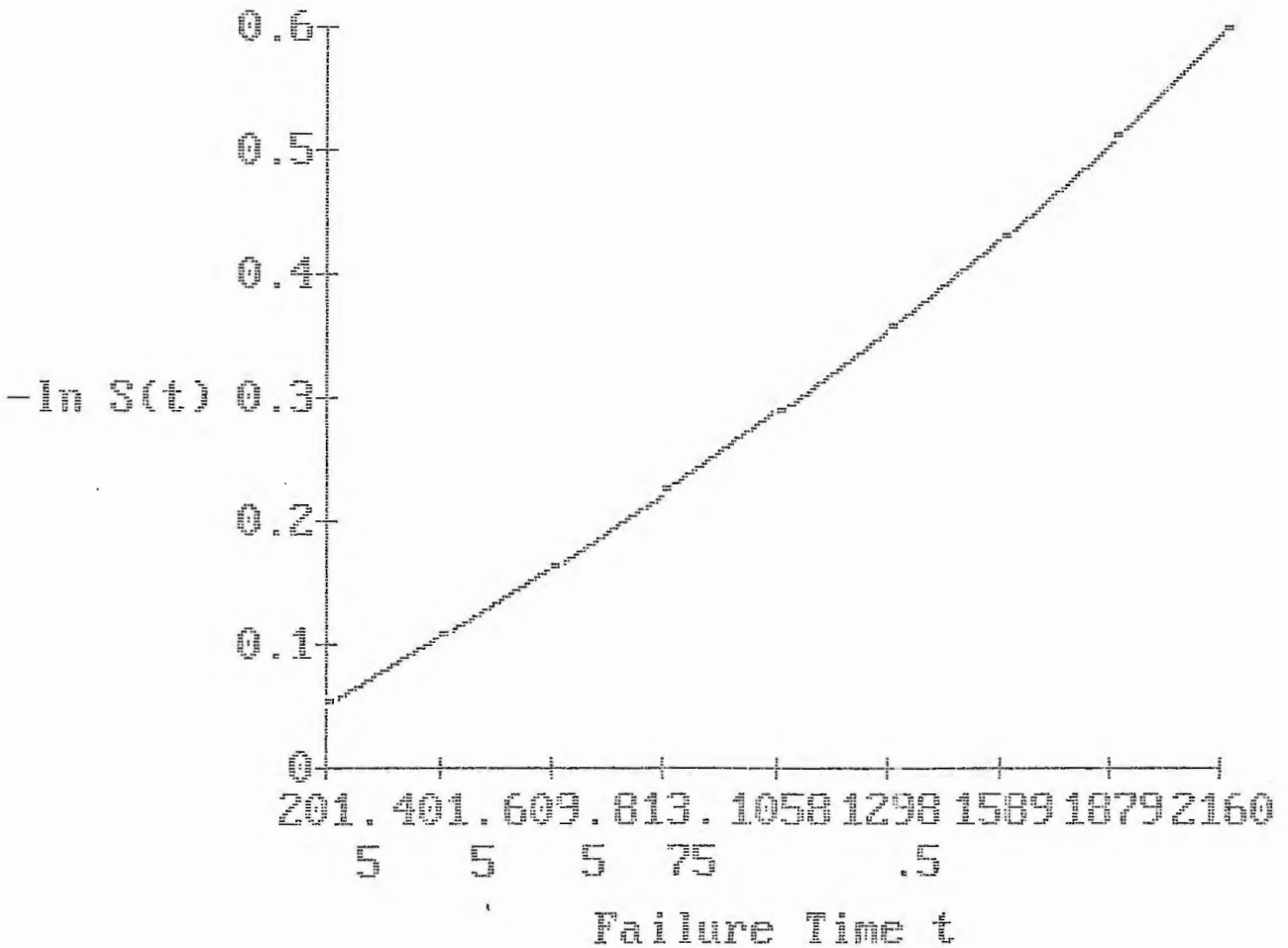




Fig.7: Graph of  $-\ln S(t)$ ,  
against Failure Time  $t$  for  
Bengal 100 Watt Bulbs



constant and a positive linear term constant were observed to be statistically significant for all the six categories of lifetimes. The coefficients of fitted lines with their standard errors are shown in table 1. The fitted lines dectate that the underlying life distributions clearly follows a two-parameter negative exponential distribution differing in the guarantee period and expected lifetimes with each other.

Table - 1: The fitted polynomials for six different categories of incandiscent electric bulbs.

Categories	The fitted polynomials
Phillips-60W	$\hat{H}(t) = - 2.5926169 + 1.87768 \times 10^{-3}t$ (0.0161456) (4.29303 x 10 <sup>-5</sup> )
Belco-60W	$\hat{H}(t) = - 0.16686420 + 3.45564 \times 10^{-4}t$ (0.0001729) (5.94709 x 10 <sup>-6</sup> )
Bengal-60W	$\hat{H}(t) = - 0.6209808 + 1.10228 \times 10^{-3}t$ (0.0041068) (2.87496 x 10 <sup>-5</sup> )
Phillips-100W	$\hat{H}(t) = - 2.311633 + 1.49006 \times 10^{-3}t$ (0.0108923) (3.16105 x 10 <sup>-5</sup> )
Belco-100W	$\hat{H}(t) = - 2.4381843 + 1.59606 \times 10^{-3}t$ (0.0123059) (3.43895 x 10 <sup>-5</sup> )
Bengal-100W	$\hat{H}(t) = - 0.0050471 + 2.76858 \times 10^{-4}t$ (4.19201 x 10 <sup>-6</sup> ) (4.59519 x 10 <sup>-6</sup> )

(\*) Figures in the parenthesis indicate standard errors of the estimate.

In each case, the coefficient of the linear term or in other words, the regression coefficient, which may be considered as scale parameter of a location-scale family of distribution, is less than unity. It implies that there is no possibility to confuse the underlying distributions as log-normal.

#### 4.3.1 Least square estimates

Since the survival function of the two parameter exponential distribution is available in closed form, it is easy to set the least square estimates of the parameters. Let

$$f(t) = \frac{1}{\theta} \exp\left(-\frac{t-m}{\theta}\right), \quad t \geq m$$

$$\Rightarrow S(t) = \exp\left[-(t-m)/\theta\right] \dots \quad (4.1)$$

Taking Ln on both sides of equation (4.1), we get,

$$\ln S(t) = -(t-m)/\theta$$

$$\text{or, } -\ln S(t) = \frac{t}{\theta} - \frac{m}{\theta}$$

$$\text{or, } -\ln S(t) = -\frac{m}{\theta} + \frac{t}{\theta} \dots \quad (4.2)$$

which can be written as a simple regression model

$$Y = a + bt + u \dots \quad (4.3)$$

where,

$$Y = -\ln S(t), \text{ dependent variable.}$$

$a = -m/\theta$ , unknown parameter

$t =$  Observed failure point in time.

$b = \frac{1}{\theta}$ , unknown parameter.

$u =$  random error observed on  $S(t)$ .

Graphs (2-7) shows the suitability of choosing a simple regression model.

The values of  $a$  and  $b$  for all the six categories of incandiscent electric bulbs are available from Table 1. Where  $\hat{\theta} = \frac{1}{b}$  and  $\hat{m} = -\hat{a} \hat{\theta}$ .

The least square estimates of the guarantee time parameter  $m$  and the scale parameter  $\theta$  are appended below for the six categories of incandiscent electric bulbs.

Table-2: Least square Estimates of Parameters.

Name of bulb	Gurantee period $\hat{m}$	Scale parameter $\hat{\theta}$
Phi-60	1380.73	532.37
Bel-60	482.90	2893.97
Ben-60	563.36	907.21
Phi-100	1551.36	671.11
Bel-100	1527.62	626.54
Ben-100	18.23	3611.95

#### 4.3.2 Maximum Likelihood Estimates

The least square method provides us only estimates of the parameter  $m$  and  $\theta$  but not the precise estimate of the standard error of the estimates. We can predict the survival probabilities at various points in time using least square estimates but these estimates fails to give us expected lifetime estimate and a confidence interval for the true expected lifetime.

Once the form of the lifetime distribution is singled out, it is easy and mostly desired to have the maximum likelihood estimates of parameters because of their optimal properties. Since our data is type- I censored, it is essential to find out the appropriate distribution function in order to obtain the maximum likelihood estimates. Let us assume the form of the probability density function of two parameter exponential model as

$$f(t, m, \theta) = \frac{1}{\theta} \exp\left[-\frac{(t-m)}{\theta}\right], \quad m \leq t \leq \infty, \quad \theta > 0 \dots (4.4)$$

The experiment was terminated at a preassigned time  $T = 2280$  hour. Here the number of items that failed before time  $T$  is a random variable which we denote by  $R$ . Let  $P(t)$  be the probability of failure before time  $T$ . Then " $R$ " has a binomial distribution,

$$P(R = r) = \binom{n}{r} p^r q^{n-r}, \quad r = 0, 1, 2, \dots, n \dots (4.5)$$

$$\begin{aligned} \text{where } p &= P(T) = 1 - \exp - (T-m) / \theta \\ \text{and } q &= 1-P = \exp - (T-m) / \theta \end{aligned} \quad \dots \quad (4.6)$$

Here the items that failed were not replaced. The data consists of the life times,  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(r)}$  of  $r$  items that failed before  $T$  and  $(n-r)$  items that survived beyond  $T$ .

Therefore,

$$P(R = r) = \binom{n}{r} \left[ 1 - \exp(t-m)/\theta \right]^r \left[ \exp - (T-m)/\theta \right]^{n-r} \dots (4.7)$$

We consider the conditional p.d.f of the failure time, given that the item has failed before time  $T$ . This is given by

$$\begin{aligned} (T/\theta) &= \frac{\exp - (t-m)/\theta}{\theta (1 - \exp - (t-m)/\theta)} \quad \dots \quad (4.8) \\ & \quad 0 \leq t \leq T \\ &= 0, \text{ otherwise} \end{aligned}$$

Thus the joint p.d.f of  $t_{(1)}, t_{(2)}, \dots, t_{(r)}$  is given by

$$\begin{aligned} g \quad t_{(1)}, t_{(2)}, \dots, t_{(r)}/\theta \\ = \frac{n!}{\theta^r} \frac{\exp - (t_i-m)/\theta}{\left[ 1 - \exp - (T-m)/\theta \right]^r} \quad \dots \quad \dots \quad (4.9) \end{aligned}$$

The likelihood of the sample is the joint p.d.f of  $t_{(1)}, t_{(2)}, \dots, t_{(r)}$  and  $r$ . Hence

$$\begin{aligned}
 L[t_{(1)}, t_{(2)}, \dots, t_{(r)}] &= g[t_{(1)}, t_{(2)}, \dots, t_{(r)} | \theta] \binom{n}{r} \rho^r q^{n-r} \\
 &= \frac{r! n!}{\theta^r r! (n-r)!} \left[ e^{-\frac{(T-m)}{\theta}} \right]^{n-r} \frac{e^{-\frac{\sum (t_i - m)}{\theta}}}{\left[ 1 - e^{-\frac{(T-m)}{\theta}} \right]^r} \left[ 1 - e^{-\frac{(T-m)}{\theta}} \right]^{\rho} \\
 &= \frac{n!}{\theta^r (n-r)!} \left[ e^{-\frac{(T-m)}{\theta}} \right]^{n-r} e^{-\frac{\sum (t_i - m)}{\theta}} \\
 &= \frac{n!}{(n-r)! \theta^r} e^{-\frac{(n-r)(T-m)}{\theta} - \frac{\sum (t_i - m)}{\theta}} \\
 &= \frac{n!}{(n-r)! \theta^r} e^{-\left[ \sum (t_i - m) + (n-r)(T-m) \right]} \dots \quad (4.10)
 \end{aligned}$$

Taking log on both sides of (4.10), we get,

$$\log L = \log C - r \log \theta - \frac{\sum (t_i - m) + (n-r)(T-m)}{\theta} \dots \quad (4.11)$$

where  $C$  is a constant.

Differentiate both sides of the above equation with respect to  $\theta$ , we get,

$$\begin{aligned}
 \frac{\partial \log L}{\partial \theta} &= -\frac{r}{\theta} + \frac{\sum (t_i - m) + (n-r)(T-m)}{\theta^2} \\
 \Rightarrow 0 &= \frac{-r\theta + \sum (t_i - m) + (n-r)(T-m)}{\theta^2} \\
 \therefore \hat{\theta} &= \left[ \sum (t_i - \bar{m}) + (n-r)(T - \bar{m}) \right] / r \\
 &= \left[ \sum t_i - r\bar{m} + (n-r)T - n\bar{m} + r\bar{m} \right] / r \dots \quad (4.12)
 \end{aligned}$$

Now we estimate  $m$ , consider  $\theta$  fixed, say  $\theta = \theta_0$  and consider the maximization of

$L [t_{(1)}, t_{(2)}, \dots, t_{(r)}, r / \theta]$  as  $m$  varies over the permissible range  $m = t_{(i)}$   $i = 1, 2, \dots, r$ . It is quite clear that

$\text{Max } L [t_{(1)}, t_{(2)}, \dots, t_{(r)}, r / \theta]$  is attained at  $m$  which minimises  $(t_{(1)} - m)$  of the denominator of equation (4.10). Thus  $\hat{m}$  is the maximum possible value of  $m$  subject to  $m = t_{(i)}$ ,  $i = 1, 2, \dots, r$ . Or,  $\hat{m} = \text{Min} (t_{(1)}, t_{(2)}, \dots, t_{(r)})$ . We note that  $\hat{m}$  is independent of  $\theta$ . Here  $\hat{m} = t_{(1)}$  is called guarantee period i.e., an item will survive at time  $t_{(1)}$ , its probability is one. Neither  $\hat{\theta}$  is unbiased for  $\theta$ , nor  $t_{(1)}$  is unbiased for  $m$  but  $\hat{\theta}$  and  $t_{(1)}$  are jointly sufficient for  $\theta$  and  $m$ .

If  $t_{(r)}$  be the  $r$ th smallest observation in a random sample of size  $n$  from the exponential distribution given in (4.4), then

$$E(t_{(r)}^{-m}) = \theta \sum_{i=1}^r \frac{1}{n-i+1}$$

$$\Rightarrow E(t_{(1)}) = m + \frac{\theta}{n} = E(\hat{m}) \dots \dots \dots (4.13)$$

Since the maximum likelihood estimators are biased, so test of significance and confidence interval based on their asymptotic variance may lead misleading conclusions. Kalbfleish and Prantice (1980) have shown that upper limit of survival probabilities based on asymptotic confidence interval exceeds unity. Lawless (1982) suggests that for type-I censored data



$$\frac{2r\hat{\theta}}{\theta} \sim \chi^2_{(2r-1)} \quad \dots \quad (4.14)$$

$$\Rightarrow E(\hat{\theta}) = \frac{2r-1}{2r} \theta$$

$$\therefore \tilde{\theta} = \frac{2r}{2r-1} \hat{\theta} \quad \dots \quad (4.15)$$

is unbiased for  $\theta$ . In a similar way,

$$\tilde{m} = \hat{m} - \frac{\tilde{\theta}}{n} = t_{(1)} - \frac{\tilde{\theta}}{n} \quad \dots \quad (4.16)$$

is an unbiased estimation of  $m$ .

$$\begin{aligned} \text{Var}(\tilde{\theta}) &= \left(\frac{2r}{2r-1}\right)^2 \text{Var}(\hat{\theta}) \\ &= \frac{2\theta^2}{2r-1} \quad \dots \quad (4.17) \end{aligned}$$

$$\begin{aligned} \text{Var}(\tilde{m}) &= \text{Var}(\hat{m}) + \frac{\text{Var}(\tilde{\theta})}{n^2} - \frac{2}{n} \text{Cov}(\hat{m}, \tilde{\theta}) \\ &= \frac{\theta^2}{n^2} + \frac{2\theta^2}{n^2(2r-1)} \\ &= \frac{\theta^2}{n^2} \left( \frac{2r+1}{2r-1} \right) \quad \dots \quad (4.18) \end{aligned}$$

The unbiased estimates of  $m$  and  $\theta$  along with their estimated standard errors are appended below in table number 3 for the six different incandescent electric bulbs under study.

Table-3 : Unbiased M.L.E of the parameters with Estimated Standard Errors

Name and power of bulb	Guarantee parameter $\tilde{m}$	Scale parameter $\tilde{\theta}$
Phillips-60W	1408.1327 (26.7443)	517.3454 (135.8615)
Belco-60W	469.4676 (148.0405)	2800.6470 (960.6140)
Bengal-60W	611.4597 (44.9228)	870.8056 (221.1848)
Phillips-100W	1547.3275 (35.1060)	673.4504 (198.5896)
Belco-100W	1601.3785 (27.0571)	519.0574 (153.0616)
Bengal-100W	26.5548 (184.9499)	3498.9035 (1200.1140)

\* Figures in the parenthesis indicate standard error of the estimates.

The mean of the distribution defined in (4.4) is given by

$$E(T) = m + \theta \quad \dots \quad (4.19)$$

Therefore, the unbiased estimated average life is given by

$$\widetilde{E}(T) = \widetilde{m} + \widetilde{\theta} \quad \dots \quad (4.20)$$

Estimated variance of the estimated mean life is given by

$$\text{Var } \tilde{E}(T) = \text{Var}(\tilde{m}) + \text{Var}(\tilde{\theta}) + 2 \text{Cov}(\tilde{m}, \tilde{\theta})$$

$\text{Var}(\tilde{m})$  and  $\text{Var}(\tilde{\theta})$  are given by (4.18) and (4.17)

$$\begin{aligned} \text{Cov}(\tilde{m}, \tilde{\theta}) &= \text{Cov} \left[ \left( \hat{m} - \frac{\hat{\theta}}{n} \right), \frac{2r}{2r-1} \hat{\theta} \right] \\ &= \text{Cov} \frac{2r}{2r-1} (m, \theta) - \frac{2r}{n(2r-1)} \frac{2r}{2r-1} \text{Var}(\hat{\theta}) \\ &= 0 - \frac{4r^2}{n(2r-1)^2} \frac{(2r-1)\theta^2}{2r^2} \\ &= - \frac{2\theta^2}{n(2r-1)} \end{aligned}$$

$$\therefore \text{Var } \tilde{E}(T) = \frac{\theta^2}{(2r-1)n^2} (2n^2 - 4n + 2r + 1) \dots \quad (4.21)$$

The distribution of  $\tilde{E}(T)$  is asymptotically normal with mean  $E(T)$  and variance defined by (4.21). Therefore,

$$Z = \frac{\tilde{E}(T) - E(T)}{\sqrt{\text{Var } \tilde{E}(T)}} \sim N(0, 1) \dots \quad (4.22)$$

The quantity  $Z$  defined in (4.22) can be used as pivotal quantity which provides confidence limits and significance test for the mean life. The unbiased estimates of mean life length along with 95% asymptotic confidence intervals for the six different brands of incandiscent electric bulbs under study are given below in table number 4.

Table-4: Unbiased Estimates of Mean Lifes.

Name and power of the bulb	Estimated Mean life $\tilde{E}(\tau)$	95% Asymptotic interval	Confidence
Phillips-60W	1925.6724 (131.6350)	1667.6677	2183.6771
Belco-60W	3270.1146 ( 923.2645)	1460.5762	5079.7130
Bengal-60W	1482.2653 (214.5892)	1061.6705	1902.8601
Phillips-100W	2220.7779 (191.6416)	1845.1604	2596.3954
Belco-100W	2120.4358 (147.7064)	1830.9363	2409.9403
Bengal-100W	3525.4583 (1153.4525)	1264.6914	5786.2252

\*Figures in the parenthesis indicate standard error of the estimates.

It is observed from Table-4 that the estimated mean life of different brands with same power are differing in a wide margin. Naturally question arises whether these observed differences in the estimated mean values are statistically significant or not. Since the asymptotic distribution of the estimated mean life are normal.

$$t = \frac{|\tilde{E}_1(\tau) - \tilde{E}_2(\tau)|}{\frac{\theta^*}{n} \sqrt{\frac{2n^2 - 4n + 2r_1 - 1}{2r_1 - 1} + \frac{2n^2 - 4n + 2r_2 + 1}{2r_2 - 1}}}$$

$$\begin{aligned}
 &= \frac{|\tilde{E}_1(\tau) - \tilde{E}_2(\tau)|}{\frac{\theta^*}{n} \sqrt{\frac{2n(n-2)}{2r_1-1} + 1 + \frac{2n(n-2)}{2r_2-1} + 1}} \\
 &= \frac{|\tilde{E}_1(\tau) - \tilde{E}_2(\tau)|}{\sqrt{n(n-2)} \frac{\theta^* \sqrt{2}}{n} \sqrt{\frac{1}{2r_1-1} + \frac{1}{2r_2-1} + \frac{1}{n(n-2)}}} \\
 &= \frac{|\tilde{E}_1(\tau) - \tilde{E}_2(\tau)|}{\theta^* \sqrt{\frac{2(n-2)}{n} \left( \frac{1}{2r_1-1} + \frac{1}{2r_2-1} \right) + \frac{2}{n}}} \text{ with } (r_1+r_2-1) \text{ d.f.} \quad (4.23)
 \end{aligned}$$

follows student's t-distribution where  $\hat{\theta} = (r_1 \hat{\theta}_1 + r_2 \hat{\theta}_2) / (r_1 + r_2 - 1)$

is the pooled estimate of  $\theta$  and is the appropriate test statistic for testing significance difference between Mean life's of two categories of same power. The calculated values of  $t$  with appropriate d.f. for pairwise difference are shown in table- 5 below.

Table-5: Calculated values of Student's t-statistic for pairwise differences in Mean life.

Power of bulb in Watt/hour	Pair compared	Degrees of freedom	Calculated value of t
60W	Phillips-Belco	23	2.25549*
	Phillips-Bengal	30	1.6890086
	Belco-Bengal	24	2.6672209*
100W	Phillips-Belco	23	0.4012851
	Phillips-Bengal	20	1.5161952
	Belco-Bengal	20	1.7130034

#### 4.4 Goodness of fit

Based on plotting procedures and polynomial fittings, we have decided that the underlying distribution of survival times of incandescent electric bulbs are two parameter exponential. Accordingly we have estimated the parameters and survival probabilities using least square and maximum likelihood methods. Now we are to ascertain whether the estimated survival probabilities are in agreement with the observed ones. If the degree of deviation is insignificant between the observed and estimated set of survival probabilities than we can conclude that our choice of model is justified and we can go into further analysis. Otherwise, we are to search for an alternative model.

The method we are using for testing goodness of fit is the likelihood ratio test. Since our censoring time is fixed, so no assumption about censoring is required. Let  $S_j$  be the probability of survival beyond  $t_j$ . The underlying hypothesis to test the goodness of fit are,

$$\begin{aligned} H_0: S_j &= S_{j0}, \quad j = 1, 2, \dots, r && \dots \quad \dots && (4.24) \\ H_1: S_j & \text{ satisfy only } 0 \leq S_j \leq 1 \end{aligned}$$

For testing the above hypothesis, the required test statistic is

$$X^2 = -2 \ln \frac{L(\hat{S}_{10}, \hat{S}_{c0}, \dots, \hat{S}_{r0})}{L(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_r)} \sim X^2_{(1)} \quad \dots \quad (4.25)$$

where  $L(\hat{S}_{10}, \hat{S}_{c0}, \dots, \hat{S}_{r0})$  is the estimated likelihood function under  $H_0$  and  $L(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_r)$  is the estimated likelihood function based on

observed survival probabilities i.e., P.E.E. of survival probabilities.

Under  $H_0$ , we have

$$S(t) = \exp\left(-\frac{t-m}{\theta}\right)$$

$$\begin{aligned} \Rightarrow L(\hat{S}_{10}, \hat{S}_{20}, \dots, \hat{S}_{r0}) &= \hat{S}(t_1) \hat{S}(t_2) \dots \hat{S}(t_r) \\ &= \exp\left(-\sum \left(\frac{t_i - \tilde{m}}{\tilde{\theta}}\right)\right) \dots \end{aligned} \quad (4.26)$$

$$\text{Similarly } L(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_r) = \prod_{j=1}^r \prod_{j=t_j < t} \left(\frac{n_j - d_j}{n_j}\right) \dots \quad (4.27)$$

Finally our test statistic becomes

$$\chi^2 = \frac{\exp\left[-\sum_{i=1}^r \left(\frac{t_i - \tilde{m}}{\tilde{\theta}}\right)\right]}{\prod_{j=1}^r \prod_{j=t_j < t} \left(\frac{n_j - d_j}{n_j}\right)} \dots \quad (4.28)$$

The values of estimated chi-squares for six different categories of incandescent electric bulbs under study are shown in Table-6.

Table-6: Values of Estimated Likelihood Ratio Statistic

Name and power of the bulb	Estimated Likelihood-Ratio	
	Using unbiased m.l. estimates	Using least square estimates
Phillips-60W	1.4844	1.0079
Belco-60W	0.8740	0.9993
Bengal-60W	1.4512	0.9989
Phillips-100W	0.9461	0.9894
Belco-100W	1.4791	1.0009
Bengal-100W	0.9350	0.9993

The theoretical value of chi-square of 5% level of significance with one degree of freedom is 3.84 which is larger than all our estimated values. This implies that the observed data is well graduated by a two parameter exponential model so that we can make inference using the estimated model.

#### 4.5 Comparison of Survival Curve

Suppose that lifetimes and censoring times are available from each of two populations, specified by

$$f(t, \mu_i, \theta_i) = \frac{1}{\theta_i} e^{-\frac{(t - \mu_i)}{\theta_i}}, \quad \mu_i \leq t \leq \infty, \quad \theta_i > 0, \\ i = 1, 2$$

we define a dummy regressor variable  $x$  that takes on values 1 or 0



according to whether an individual comes from the first or the second population so that the covariate vector  $X = (1 \ 0)$  if the individual belongs to the first population when  $X = (0 \ 1)$  if the individual belongs to the second population. Now we can write the hazard function as

$$h(t|x) = h_0(t) e^{x\beta} = \frac{1}{\theta} e^{x\beta}$$

for the combined population. Let  $S_{1i}$  be the conditional probability of surviving  $t_{(i)}$  for individuals from the first population and  $S_{2i}$  be the corresponding survival probability for individuals from the second population. The hazard function for the two distributions are

$$h_1(t) = h_0(t)e^{\beta} \quad \text{and} \quad h_2(t) = h_0(t)$$

The two distributions are identical iff  $\beta = 0$ . A test of  $\beta = 0$  is equivalent to the test  $S_1(t) = S_2(t)$ .

Let  $n_i$  be the total number of individuals at risk just prior to  $t_{(i)}$  from the combined sample and  $n_{1i}$  and  $n_{2i}$  be the corresponding number for the first and second population, so that  $n_i = n_{1i} + n_{2i}$ . Similarly let  $d_{1i}$  and  $d_{2i}$  be the number of failures from the first and the second population respectively at  $t_{(i)}$  such that  $d_i = d_{1i} + d_{2i}$  be the total number of failure at  $t_{(i)}$ .

With the combined sample from the two populations and the dummy regressor variable  $x$ , the score vector of exponential ordered score

test (sometimes referred to as Savage or Log-rank test) with censored data is given by

$$U(o) = \sum_{i=1}^k \left( d_{1i} - \frac{d_i n_{1i}}{n_i} \right), \quad i = 1 \text{ or } 2$$

The information quantity (variance of the difference) is given by

$$I(o) = \sum_{i=1}^k \left( \frac{d_i(n_i - d_i) n_{1i} n_{2i}}{n_i^2 (n_i - 1)} \right)$$

A test of  $H_0: \beta = 0$  (equivalently,  $S_1(t) = S_2(t)$ ) can be based on either

$$Z = \frac{U(o)}{[I(o)]^{1/2}} \sim N(0, 1)$$

Large absolute values of  $Z$  provide evidence against equality of the two survival curves. Alternatively,

$$Z^2 = \frac{[U(o)]^2}{I(o)} \sim \chi^2(1)$$

The later is sometimes referred to as Mantel-Haenszel (1959) test. The calculated values of  $Z^2$  for pairwise comparison are shown in table 7.

Table-7 : Pairwise Comparison of Survival Curves for different Power Categories/Brands.

Power	Comparison Pair	Calculated $Z^2$
60	Phillips-Belco	5.2755*
	Phillips-Bengal	3.0671
	Belco-Bengal	6.1918*
100	Phillips-Belco	0.0164
	Phillips-Bengal	0.1835
	Belco-Bengal	0.1852

\*Significant at 5% level.

## CHAPTER FIVE

### STUDY OF COST-BENEFIT FACTORS OF DIFFERENT BRANDS OF INCANDISCENT ELECTRIC BULBS

#### 5.1 Introduction

In Chapter Four, we have studied the reliability parameters, that is, guarantee period, intensity of failure and average life time of different brands of incandiscent electric bulbs by power categories under investigation. It has been discussed at length in Chapter One that the overall quality of an incandiscent electric bulb depends not only on reliability parameters but also on the benefit factor, that is, the intensity of light and the cost factor, that is, the amount of power that it consumes per unit time to unit the light. Hence, only a larger mean life time or a larger guarantee period or a smaller intensity of failure of one branch can not assure that it is better than any other branch unless we study the cost benefit factors and calculate how much we are to pay for per unit of light. This chapter is aimed to study these cost benefit factors of different brands of incandiscent electric bulbs under investigation.

#### 5.2 Measurement of Intensity of Light

Following instruments are essential to measure the intensity of light:

- 1) Lux meter : It is an instrument which can measure the intensity of light directly in the unit of lux. It has a reflector. The light of the glowing bulb from unit distance falls directly on the reflection and the lux meter in this reflector read the intensity of light in lux which one lux = 0.093
- 2) Variac or Variable Stabilizer: This instrument can stabilize the voltage of electricity flow in the range of 0-240 volts and any desired voltage can be used to observe the variation in the intensity of light. But the variac itself cannot read the voltage. Since voltage in a variable range can be stabilized, the instrument is called variac for this reason.
- 3) Avometer or Multimeter : It is an electric meter which can measure the flow of electricity through a conductor, the voltage of electricity and many other electrical parameters. We have used an avometer to measure the line voltage through the variac since variac itself cannot measure the voltage.
- 4) Incandiscent electric bulbs under study.
- 5) A meter scale.

Measurement of intensity of light is an electro-physical experiment which require a standard dark room. Department of statistics has neither dark room nor these appliances to conduct the experiment. We have converted a room of the department of Statistics temporarily into dark room

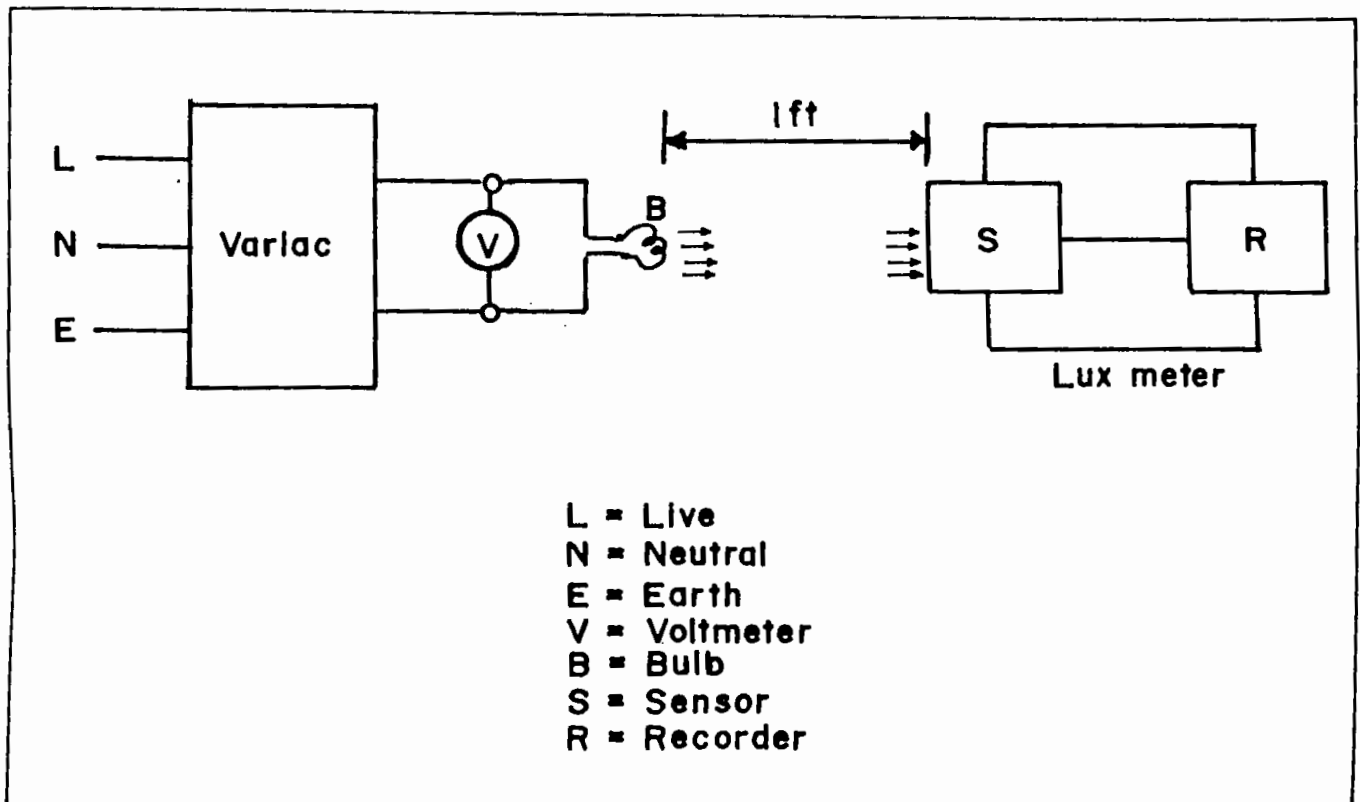


Fig.8. Circuit Diagram For Measuring Intensity of Light .

using black screens, while the Department of Applied Physics and Electronics helped us with instruments and technical know-how to run the experiment. The outline of the experimental setup is shown in circuit diagram (Fig. 8).

The stabilizer was connected to the electric line directly. The bulb under experiment and the avometer was connected to the line through the stabilizer. The distance of the filament of the bulb to the centre of the reflector of the lux meter was fixed at one feet. The voltage was fixed first by adjusting the variac and then the bulb was made on to measure the intensity of light for that voltage. The lux meter could not give any reading for intensity of light below 60 volt. The reason is that a certain amount of power is required first to heat the filament, only after which light can be emitted. Intensity of light was measured for three bulbs of each category and the experimental results are appended in Appendix V.

### 5.3 Voltage-Intensity Relationship

According to Ohm's law, the ratio of potential difference  $V$  between any two points on a conductor to the current  $C$  flowing between them, is constant, provided the temperature of the conductor does not change. In other words,

$$\frac{V}{C} = \text{Constant}$$

$$\text{or, } V = RC$$

where  $R$  is the resistance of the conductor between the points considered. It simply means that provided  $R$  remains constant, current is directly proportional to the potential difference across the ends of a conductor. In case of an incandescent electric bulb, the filament of the bulb is the conductor in question which becomes red hot and finally glows to emit light when current is passed through it. The amount of current passing through the filament is proportional to the voltage of electric line, no doubt, but not directly. The reason is that a portion of the current passing through the filament is converted to heat which changes the resistance of the filament. It implies that the amount of light emitting by an incandescent electric bulb in unit time is related to the line voltage but the nature of the relationship is not exactly known. Plots of the measured intensity of light against line voltage in Figures 9-14 indicate a polynomial relationship between intensity of light and the line voltage for all the bulbs under study but the degree of polynomials may vary from one another in degree or in parameters and hence in expected intensity for given average line voltage. We have fitted third and four degree polynomials to every set of data in the range of 60-240 voltage. The coefficients are found significant upto third degree. When four degree is included the computer accepts four degree but regrets third degree. If we replace third degree by four degree, the improvement in the sum of squares regression is insignificant. This recommends a third degree polynomial to explain the voltage intensity relationship in the range of 60-240 volt for all the brands of incandescent electric bulbs. In other words,

$$E (I/60 \leq V \leq 240) = b_0 + b_1 V + b_2 V^2 + b_3 V^3 \quad \dots \quad (5.1)$$



Fig. 9: Graph of Intensity of Light against Voltage for Phillips 60 Watt Bulbs

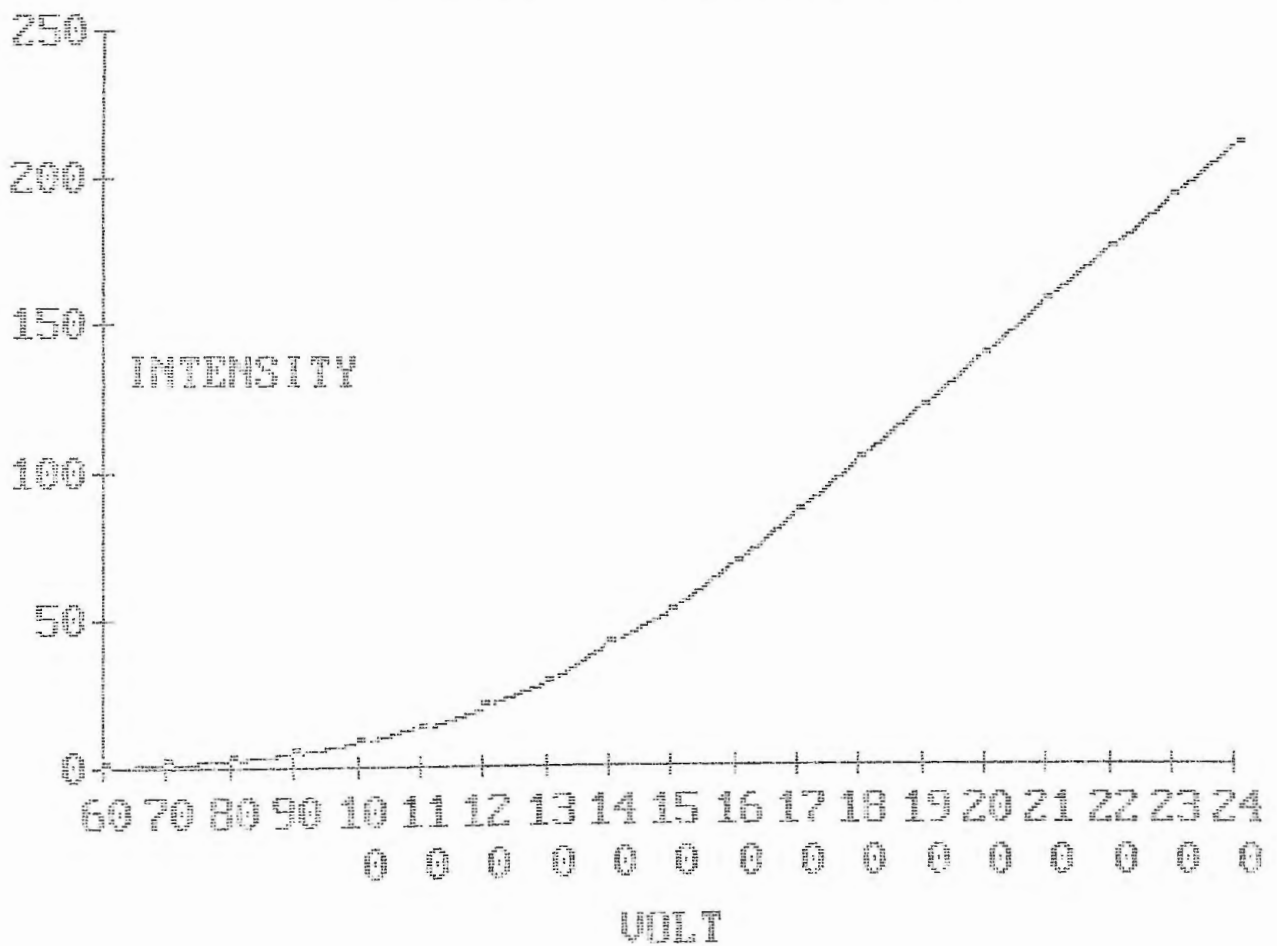


Fig.10: Graph of Intensity of Light against Voltage for Belco 60 Watt Bulbs

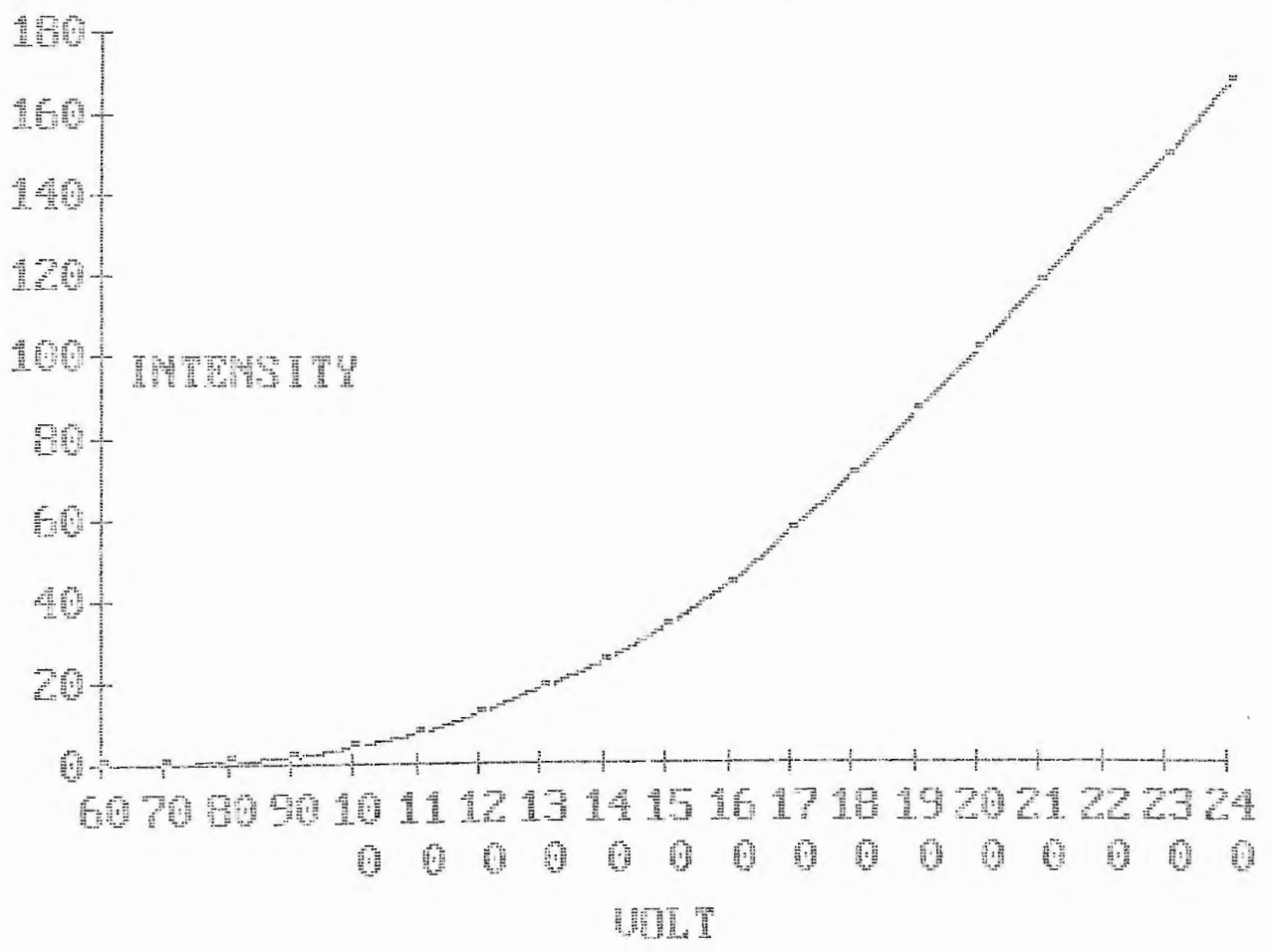


Fig. 11: Graph of Intensity of Light against Voltage for Bengal 60 Watt Bulbs

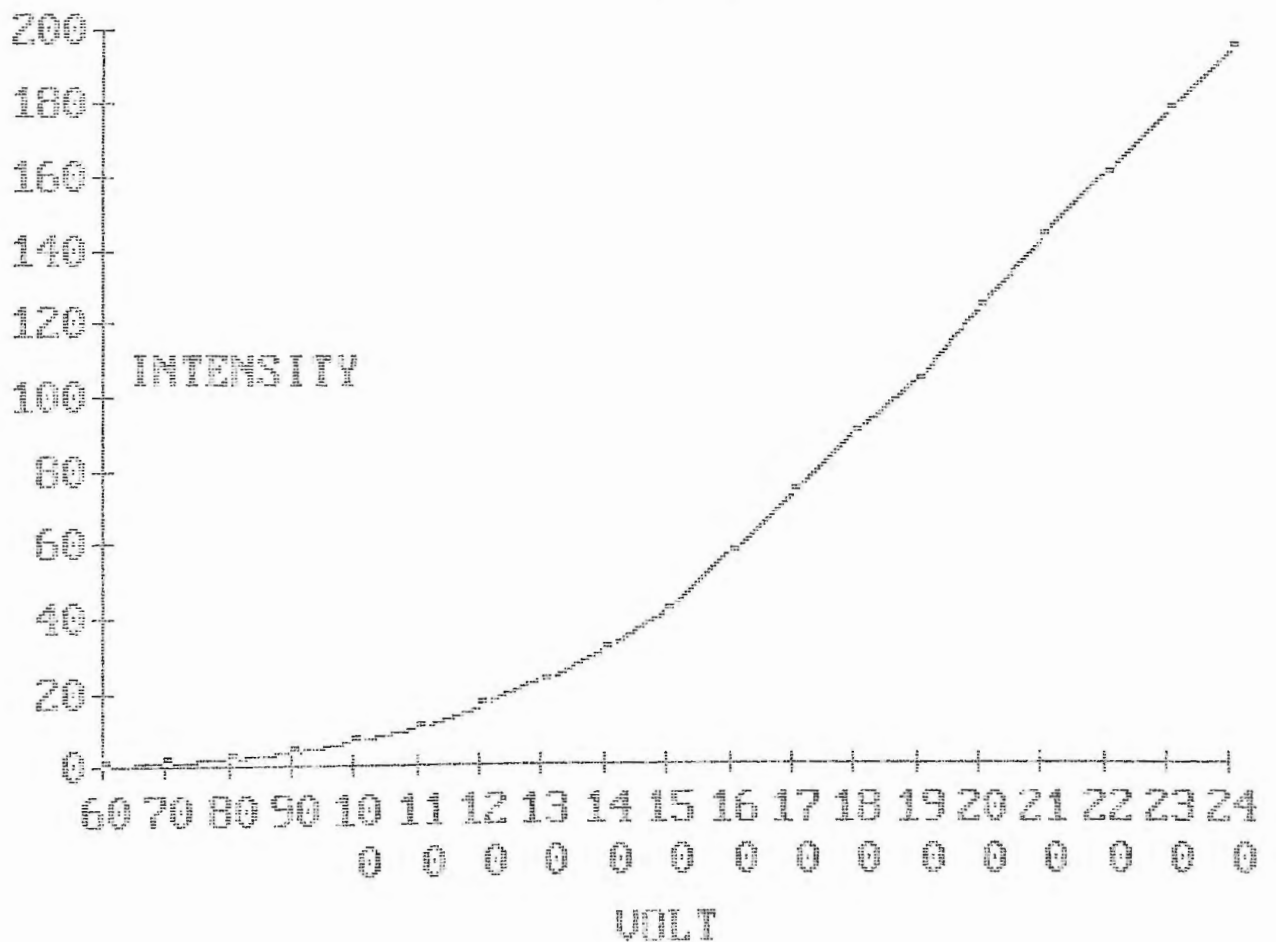


Fig.12: Graph of Intensity of Light against Voltage for Phillips 100 Watt Bulbs

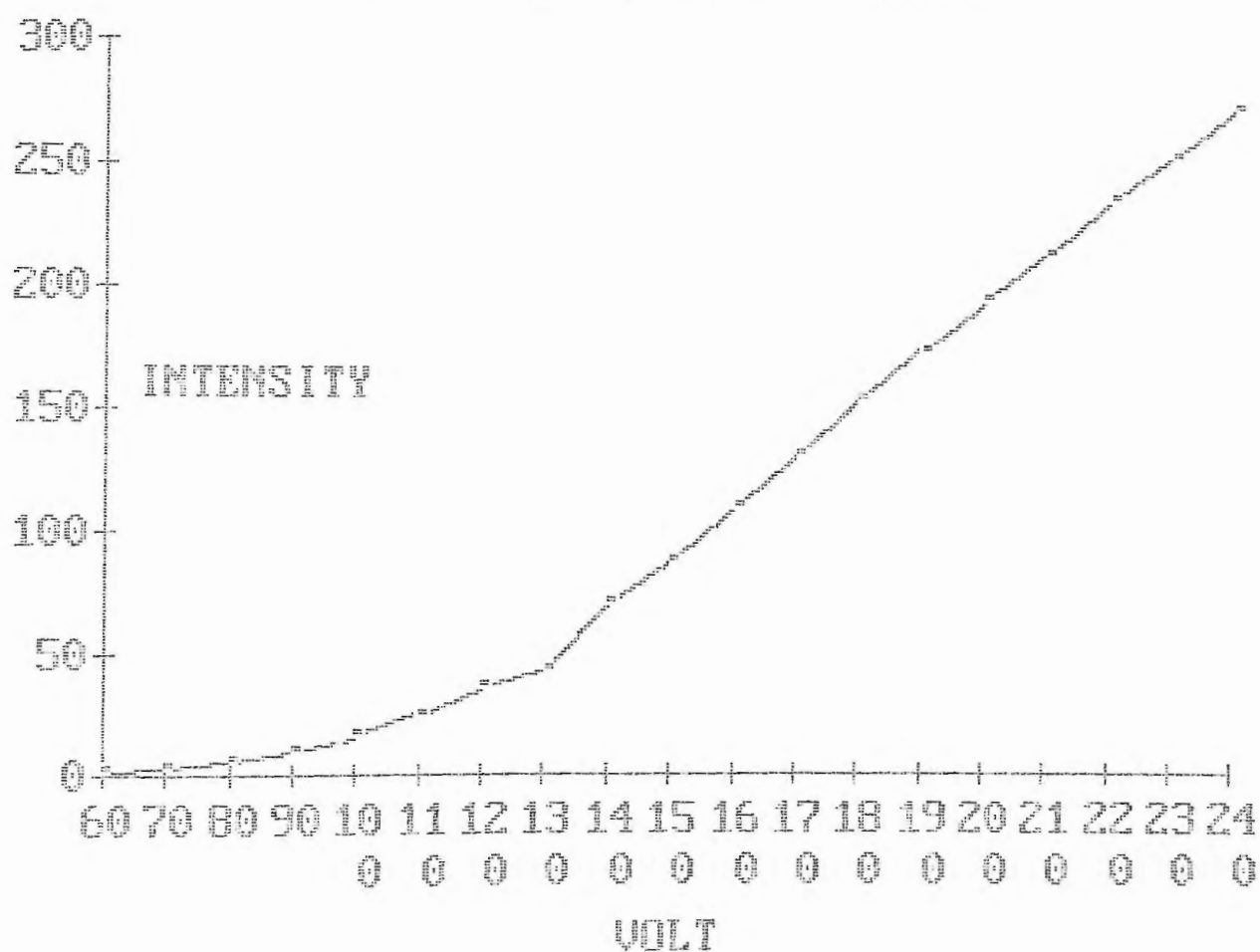


Fig-13: Graph of Intensity of Light against Voltage for Belco 100 watt Bulbs

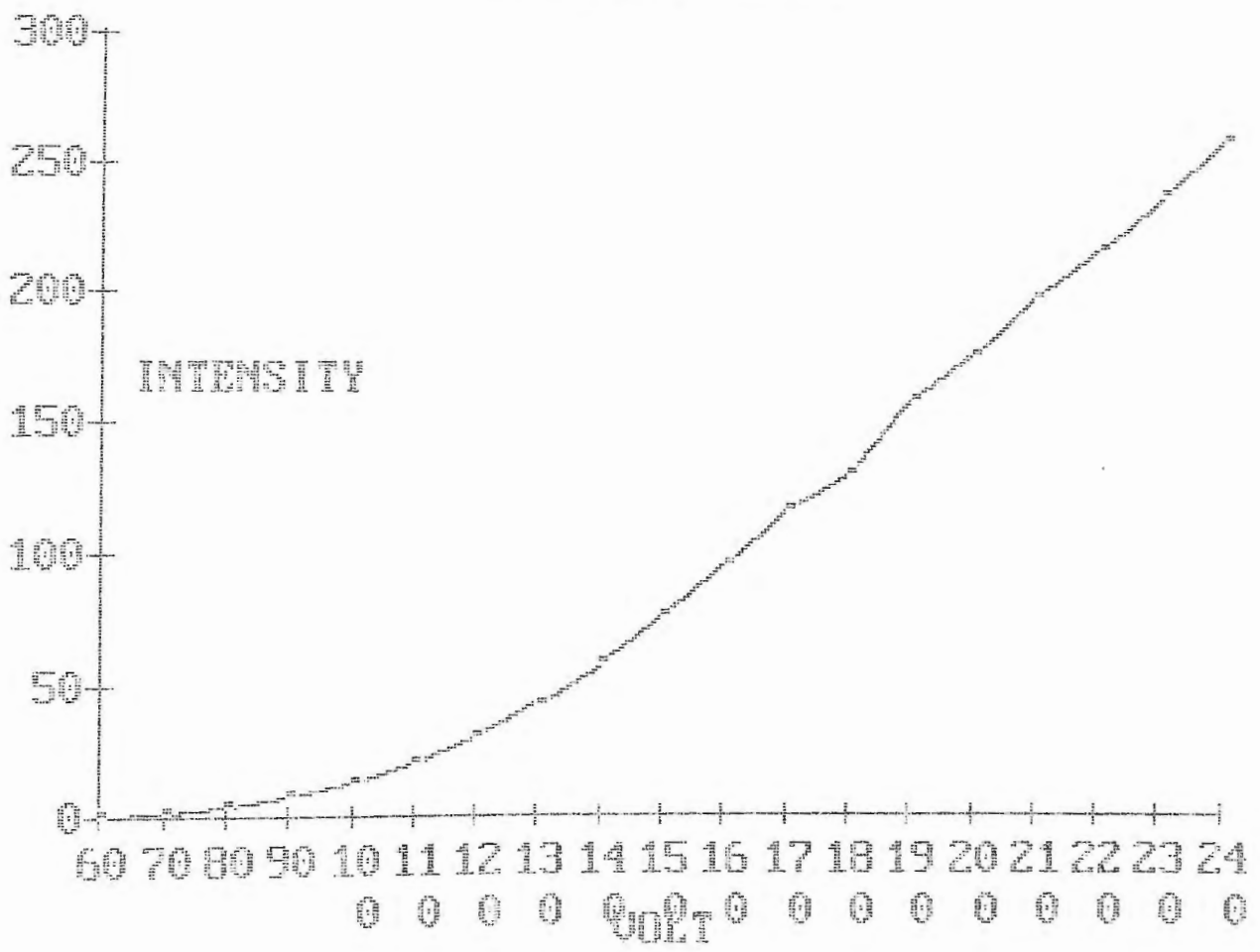


Fig 14: Graph of Intensity of Light against Voltage for Bengal 100 watt Bulbs

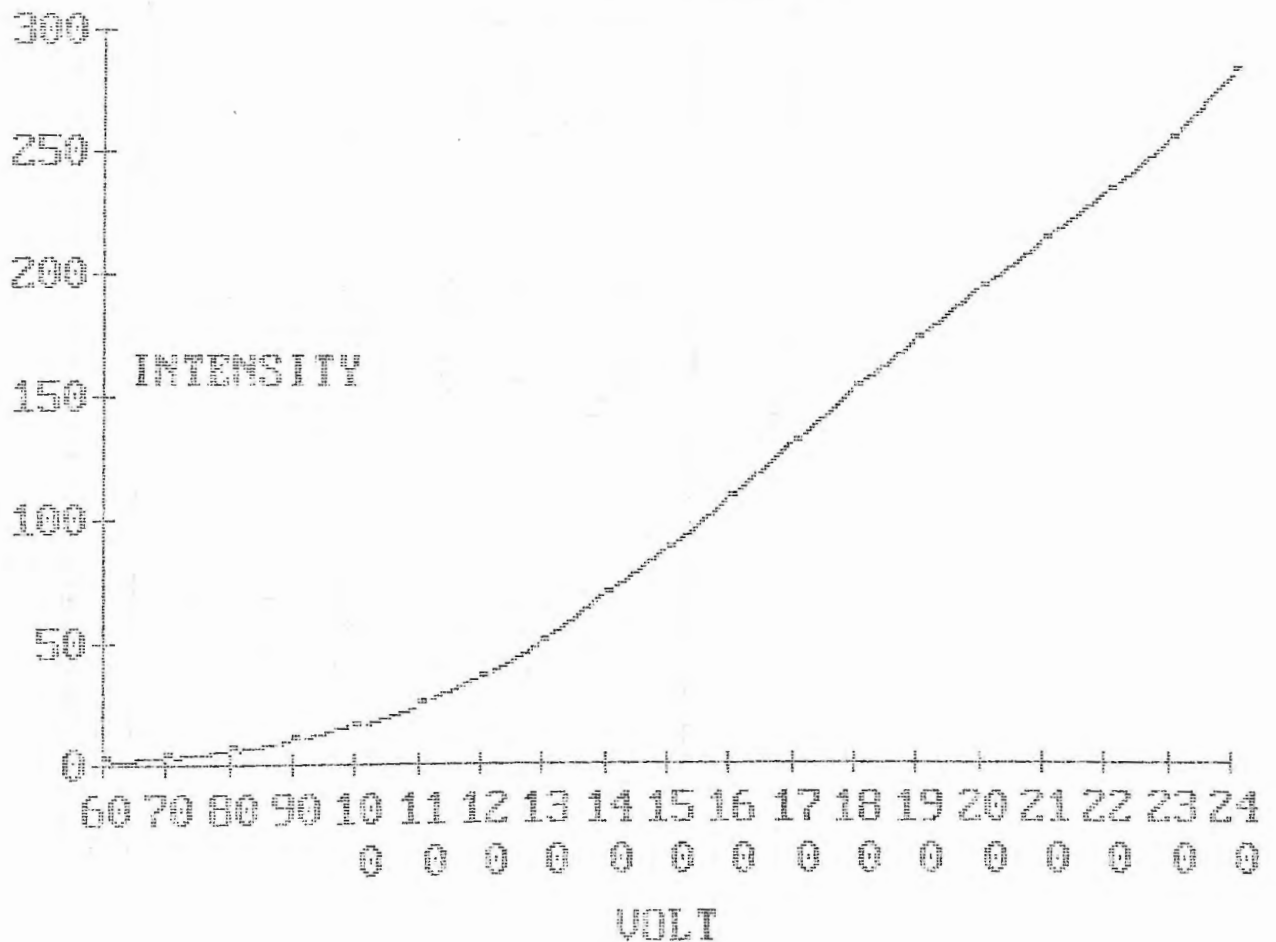


Table 8: Fitted Polynomials of Voltage-Intensity Relationship.

Power	Brand	$b_0$	$b_1$	$b_2$	$b_3$	$R^2$
60	Phillips	90.17227 (10.97824)	-2.49267 (0.25411)	0.01937 (0.001803)	-0.00002856 (0.00000398)	0.99922 (2.20748)
	Belco	53.60156 (9.40419)	-1.41526 (0.21768)	0.00991465 (0.00154439)	-0.00000837 (0.00000341)	0.99905 (1.89098)
	Bengal	81.52971 (13.29776)	-2.15650 (0.30781)	0.01581 (0.00218381)	-0.00001997 (0.00000482)	0.99865 (2.67389)
100	Phillips	127.12703 (15.98724)	-3.72416 (0.37006)	0.03159 (0.00262548)	-0.00005644 (0.000005802)	0.99903 (3.21468)
	Belco	99.64807 (23.19753)	-2.87951 (0.53696)	0.02364 (0.00380958)	-0.00003653 (0.00000842)	0.99775 (4.66451)
	Bengal	110.26486 (13.99016)	-3.28646 (0.32383)	0.02813 (0.00229752)	-0.00004794 (0.00000508)	0.00027 (2.81311)

Figures in the parenthesis indicate standard error of estimates.

where  $I$  is the amount of light emitting per second by an incandescent electric bulb,  $V$  is the line voltage and  $b_0$ ,  $b_1$ ,  $b_2$  and  $b_3$  are constants. The fitted polynomials in the unit of lux are shown in Table 8.

Though the voltage-intensity relationship in the range of (60-240) volt is found to be a third degree polynomial, actually the line voltage varies in the range of (190-240) volts in general. In this range, the voltage-intensity relationship is found to be a linear one, that is,

$$E (I/190 \leq V \leq 240) = a + bV \quad \dots \quad \dots \quad (5.2)$$

The fitted linear relationship in the unit of lux for different brands of incandescent electric bulbs are shown in Table 9.

Table 9: Fitted Voltage-Intensity Linear Relationship

Power	Brand	a	b	R <sup>2</sup>
60	Phillips	-218.43429 (2.43699)	1.79086 (0.01130)	0.99984 (0.47268)
	Belco	-219.48214 (4.47158)	1.60893 (0.02073)	0.99934 (0.86731)
	Bengal	-236.83929 (6.94317)	1.80536 (0.03219)	0.99873 (1.34671)
100	Phillips	-202.51619 (4.37227)	1.98457 (0.02027)	0.99958 (0.84805)
	Belco	-255.53143 (42.05389)	2.16829 (0.19499)	0.96867 (8.15682)
	Bengal	-234.24571 (14.94205)	2.14114 (0.06928)	0.99583 (2.89818)

Figures in the parenthesis indicate standard error of estimates.



#### 5.4 Choice of the Model for Predicting Intensities

The average line voltage estimated during the period of experiment is found to be ( $V = 215.00$ ) and the estimated intensity at this average line voltage as per fitted polynomial, and linear relationship for different brands are shown in Table 10.

Table 10: Estimated Intensity at Average Line Voltage  $V = 215$  of Different Brands for the Two Models

Power	Brand	Estimated polynomial	Intensity in Lux Linear
60	Phillips	165.7457 (2.1318)	166.6006 (0.1909)
	Belco	124.4192 (0.7880)	126.4378 (0.3614)
	Bengal	150.2102 (1.1045)	151.3131 (0.5563)
100	Phillips	225.6760 (1.3356)	224.1664 (0.3523)
	Belco	210.2526 (1.8970)	210.6509 (3.3173)
	Bengal	227.4799 (1.1476)	226.0994 (1.1823)

Figures in the parenthesis indicate standard error of the estimates.

Estimated intensity displayed in Table 10 reflects that at the average line voltage  $V = 215$ , the third degree polynomial and the linear model reproduce almost identical results. Moreover, standard error of

estimates are, in general, higher for the polynomials than the linear fittings. This influences us to consider the linear model (5.2) instead of the polynomial (5.1) for estimating the average intensity at the average line voltage. It is to be noted that the estimated value of  $a$  for all the brands under study are negative. This is due to the fact that a certain portion of the total power used are converted into heat and the parameter  $a$  may be considered as a function of that portion of power in absolute value. The parameter  $b$  may be interpreted as the rate of change of intensity for unit change in voltage beyond  $a$ .

### 5.5 Comparison of Parameters for different Brands

We are paying for the power to get light only. The portion of the power consumed in heating purpose may be considered as wastage. The less is the amount of wastage, the higher is the amount of benefit which pin points the quality of a bulb. Naturally, one may raise the question, whether or not, the amount of wastage of power is differing significantly between any two brands of same power under study. This is equivalent to a pair-wise of test for significance difference of  $a$ . Again a higher rate of change in the intensity of light for unit change in line voltage also indicates higher benefit and quality of a bulb in question. So, we are to investigate, whether or not, the rates i.e. values of  $b$  for different brands of same power under study are differing significantly.

It is difficult to conduct a precise test for significant difference between different brands for these parameters using separate models for each brand as is shown in equation (5.2) because we are lacking of distributional concepts. Rather, we may use a single linear model combining two or three brands together using dummy variables to perform these tests under ordinary least square set up.

For the  $j$ th branch of  $i$ th power category, we may write the voltage-intensity relationship in the range of (190-240) volt as

$$E(I_{ji}/V) = a_{ji} + b_{ji}V \quad ; \quad j = 1, 2, 3; \quad i = 1, 2 \quad (5.3)$$

For each  $i$ , the three equations in (5.3) may be combined together as

$$E(I/V) = a_1 + (a_2 - a_1) D_2 + (a_3 - a_1) D_3 + b_1 V + (b_2 - b_1) D_2 V + (b_3 - b_1) D_3 V \quad \dots \quad (5.4A)$$

$$E(I/V) = a_2 + (a_3 - a_2) D_3 + b_2 V + (b_3 - b_2) D_3 V \quad \dots \quad (5.4B)$$

$$\text{where } D_i = \begin{cases} 1 & \text{for an observation of } I \text{ in } i\text{th brand} \\ 0 & \text{otherwise} \end{cases}$$

Now testing the significance of  $D_2$  in (5.4) is equivalent to testing the difference of  $(a_2 - a_1)$  in (5.3). Similar is the case with  $D_3$ ,  $D_2 V$ ,  $D_3 V$  etc. in (5.4).

Pairwise comparison of different brands by power as per model (5.4A) and (5.4B) are shown in Table 11.

Table 11 : Pair-wise Comparison of Parameters

Power	Brands	d.f.	Difference in intercepts 'a'	Difference in scale parameter 'b'
60	Belco-Phillips	12	-1.047850 (5.092538)	-0.181930*** -(0.023610)
	Phillips-Bengal	12	17.405000* (7.358470)	-0.014500 (0.034116)
	Belco-Bengal	8	17.357150 (8.258489)	-0.196430*** (0.038287)
100	Belco-Phillips	12	-53.015240 (42.280568)	0.183720 (0.196041)
	Bengal-Phillips	12	-31.729520 (15.568610)	0.156570 (0.072194)
	Bengal-Belco	8	21.285720 (44.629525)	-0.02715 (0.206932)

Figures in the parenthesis indicate standard error of the difference

\* P = 0.05, \*\*\* P  $\geq$  0.0001

It is evident from Table 11 that the scale parameters of 60 watt Phillips and Bengal are greater than that of Belco while intercept parameter of Phillips is greater than that of Bengal. This apparent difference may not suggest that intensity at average line voltage may differ significantly among the brands under study. It is interesting that in case of 100 watt, the brand do not differ significantly from one another either in intercept parameter or in scale parameter still

they may differ in intensity at the average line voltage. Pair-wise comparison of intensity at average line voltage for different power are shown in Table 12.


Table 12: Pair-wise Comparison of Intensity of Light of Different Brands by Power at the Average Line Voltage  $V = 215$

Power	Brands	Difference in intensity in Lux	S.E. of the Difference	d.f.	Value of t
60	Phillips-Belco	40.16280	0.40878	8	98.2504***
	Phillips-Bengal	15.2875	0.58813	8	25.9934***
	Bengal-Belco	24.8753	0.66338	8	37.4978***
100	Phillips-Belco	13.51544	3.33593	8	4.0515***
	Phillips-Bengal	1.93303	1.23370	8	1.5669
	Bengal-Belco	15.44847	3.52167	8	4.3867***

\*\*\* P 0.001

It is clear from Table 12 that in case of 60 watt bulbs, Phillips has the highest intensity of light which differ significantly from both Bengal and Belco, while Bengal has the second highest intensity which differ significantly from Belco at the average line voltage  $V = 215$ . Of the 100 Watt bulbs, intensity of both Phillips and Bengal are significantly higher than Belco while Phillips and Bengal do not differ significantly at the average line voltage  $V = 215$ .

## 5.6 Measurement of Power Consumption

Consumption of power is the cost factor of an incandiscent electric bulb. Unless we measure it, cost-benefit analysis is not possible. The power may be defined in various way but the simplest one to serve our purpose is the product of the line voltage and the amount of current passing through the filament of the bulb in question in one hour is the amount of power that the bulb consumes at the given line voltage. At a given line voltage  $V$  if  $C$   be the amount of current passing through the filament of the bulb under investigation in one hour, then the power consumption of the bulb at line voltage  $V$  is

$$\text{Power} = CV \text{ Watt/hour} \quad \dots \quad \dots \quad (5.5)$$

Generally we use the tr - Killo-watt/hour. If we divide the Equation (5.5) by one thousand we get power in Killo-watt/hour. To measure the power consumption, actually we need only to measure the amount of current passing through the filament of a bulb at different line voltage and establish a suitable voltage-current relationship to estimate the power consumption at the average line-voltage. This would also help us in estimating the total power consumption during the life-time of a bulb.

We have used the following instruments in measuring the amount of current passing through the filament of a bulb:

- 1) Voltmeter : It is an electrical meter which can measure the voltage of electricity.

- 2) Avometer : It is also an electrical meter which has multiple use in electrical technology. We have used the meter to measure the amount of current passing through the filament of a bulb.
- 3) Variac or Variable Stabilizer : It is an instrument which can control the voltage fluctuation and can be used at a variable range of voltage of the electricity.
- 4) Incandiscent electric bulbs under study.

The experimental set up is shown in circuit-diagram (Fig. 15). The stabilizer connected to the electric line directly. The voltmeter and the bulb under study was connected to the stabilizer in series while the avometer was connected to the bulb in parallel position. First we have fixed the voltage of the line using the variac and then we switch on the bulb to get reading on the avometer which gives the amount of current passing through the filament of the bulb under investigation in ampere/hour. Experiment was done on three bulb of each category chosen at random. The experimental results are given in Appendix VI.

### 5.7 Voltage-Current Relationship

We have discussed in Section 5.3 that there is a functional relationship between line voltage  $V$ , the resistance of the filament  $R$  and the amount of current passing through the filament  $C$ , that is, under

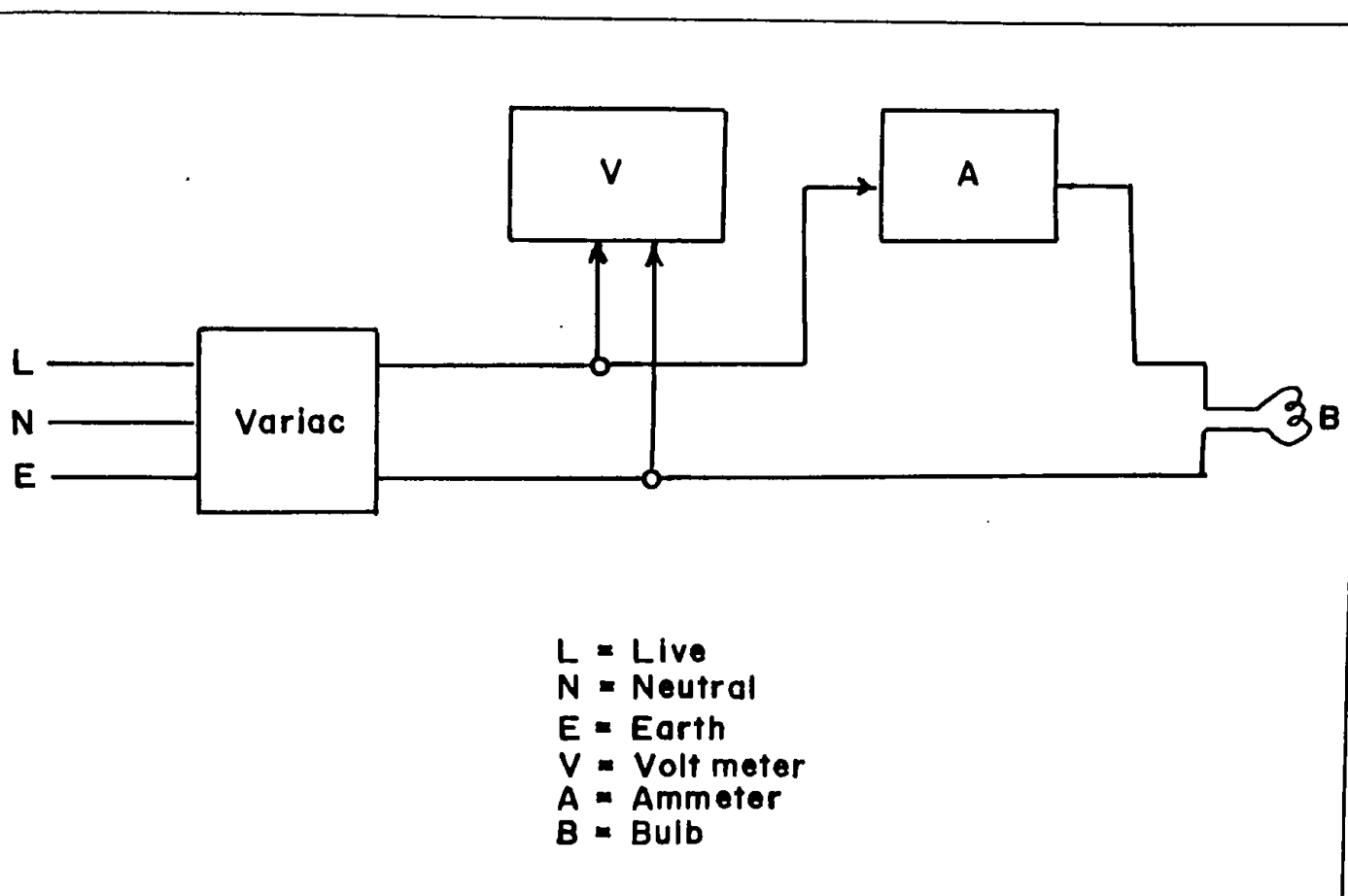


Fig.15. Circuit Diagram For Measuring Current Consumption .



constant resistance,

$$V = R C$$

$$C = \frac{V}{R}$$

But resistance increases as heat increases due to increase in line voltage. Hence the linear relationship is disturbed. The fitted relationship between line voltage and current is found to be a third degree polynomial in the range of 0.240 volt for all the brand of electric bulbs under study. In other words, what we have found is

$$E (C/0 < V < 240) = \beta_1 V + \beta_2 V^2 + \beta_3 V^3 \dots \quad (5.6)$$

where  $C$  is the amount of current passing through the filament of the bulb per hour in ampere unit. The fitted polynomials in the unit of ampere/hour are shown in Table 13. Actually, the line voltage varies between 190-240 volts in general. Within this range, the current-voltage relationship is found to be linear one, that is,

$$E (C/190 \leq V \leq 240) = \alpha + \beta V \dots \quad (5.7)$$

Table 14 delineates the fitted linear relationship in the unit of ampere/hour for different brands of incandiscent electric bulbs under study.

constant resistance,

$$V = R C$$

$$C = \frac{V}{R}$$

But resistance increases as heat increases due to increase in line voltage. Hence the linear relationship is disturbed. The fitted relationship between line voltage and current is found to be a third degree polynomial in the range of 0.240 volt for all the brand of electric bulbs under study. In other words, what we have found is

$$E (C/0 < V < 240) = \beta_1 V + \beta_2 V^2 + \beta_3 V^3 \dots \quad (5.6)$$

where C is the amount of current passing through the filament of the bulb per hour in ampere unit. The fitted polynomials in the unit of ampere/hour are shown in Table 13. Actually, the line voltage varies between 190-240 volts in general. Within this range, the current-voltage relationship is found to be linear one, that is,

$$E (C/190 \leq V \leq 240) = \alpha + \beta V \dots \quad (5.7)$$

Table 14 delineates the fitted linear relationship in the unit of ampere/hour for different brands of incandiscent electric bulbs under study.

Table 13: Fitted Polynomials of Voltage-Current Relationship.

Power	Brand	$\beta_1$	$\beta_2$	$\beta_3$	$R^2$
60	Phillips	2.704899-03 (2.54179-04)	-1.51944-05 (3.04899-06)	3.4476-08 (9.20825-09)	0.98763 (0.01045)
	Belco	2.62301-03 (2.53282-04)	-1.53421-05 (3.03824-06)	3.556567-08 (9.17577-09)	0.98603 (0.01042)
	Bengal	2.73310-03 (2.56637-04)	-1.64582-05 (3.07848-06)	3.824162-08 (9.29730-09)	0.98564 (0.01056)
100	Phillips	5.136097-03 (3.91722-04)	-3.09801-05 (4.69889-06)	7.090103-08 (1.41911-08)	1.99005 (0.01611)
	Belco	4.59314-03 (4.34015-04)	-2.69217-05 (5.20622-06)	6.108107-08 (1.57233-08)	0.98583 (0.01785)
	Bengal	4.67603-03 (3.90738-04)	-2.72077-05 (4.68709-06)	5.944091-08 (1.41555-08)	0.98814 (0.01607)

Figures in the parenthesis indicate standard error of estimates.

Table 14: Fitted Voltage-Current Linear Relationship

Power	Brand	$\alpha$	$\beta$	$R^2$
60	Phillips	0.10743 (1.42843-03)	5.808571-04 (6.62299-06)	0.99948 (2.77059-04)
	Belco	0.09365 (1.25685-03)	5.788571-04 (5.82745-06)	0.99959 (2.437798-04)
	Bengal	0.09728 (1.55568-03)	5.537143-04 (7.21299-06)	0.99932 (3.01741-04)
100	Phillips	0.18249 (1.95450-03)	9.502857-04 (9.06214-06)	0.99964 (3.790967-04)
	Belco	0.17662 (8.89438-04)	8.928571-04 (4.11393-06)	0.99991 (1.725164-04)
	Bengal	0.20405 (9.05347-04)	7.028571-04 (4.01328-06)	0.99981 (1.836253-04)

Figures in the parenthesis indicate standard error of the estimates.

### 5.8 Choice of the Model for Predicting Flow of Electricity

It is evident from Section 5.7 that the voltage-current relationship in the range of (0-240) volt is a third degree polynomial whereas in the range of (190-240) volt it is linear. The actual line voltage varies in the range of (190-240) volt in general. We are to investigate whether the amount of current estimated from the two models differ significantly at the estimated average line voltage  $\hat{V} = 215$ . The estimated current at the average line voltage as per fitted polynomial and linear relationship for different brands of bulbs under study are shown in Table 15.

Table 15: Estimated Flow of Current at Average Line Voltage

Power	Brand	Estimated Flow of Current in Ampere/Hour	
		Polynomial	Linear
60	Phillips	0.2218 (2.54197-04)	0.2323 (1.13132-04)
	Belco	0.2082 (2.53300-04)	0.2181 (9.95446-05)
	Bengal	0.2069 (2.56655-04)	0.2163 (1.23214-04)
100	Phillips	0.3768 (3.91750-04)	0.3868 (1.54800-04)
	Belco	0.3510 (4.34046-04)	0.3685 (7.04321-05)
	Bengal	0.3384 (3.90766-04)	0.3551 (2.74105-04)

Figures in the parenthesis indicate standard error of the estimates.

An eye view through the Table 15 reflects that the linear model slightly over estimates the flow of current at the average line voltage but the difference is within the permissible limit of 5%. The linear model calibrates the data at the extreme point of  $V = 240$  whereas, the polynomial over estimates at this extreme point. Moreover, in almost all cases, standard error of estimates are higher for polynomial than the linear model. Considering all those facts, a linear model is advantageous over the polynomial to predict voltage-current relationship in the range of (190-240) volts.

#### 5.9 Comparison of Power Consumption for Different Brands

Once we know the voltage-current relationship, we can utilize the formula (5.5) to estimate the power consumption of an electric bulb at any desired line voltage. As per formula (5.5), we have

$$\begin{aligned} \text{Power} &= V C \text{ Watt/Hour} \\ &= \frac{V \cdot C}{1000} \text{ k.w./hour} \end{aligned}$$

The estimated flow of current in ampere/hour is given in Table 15 at the average line voltage  $V = 215$ . The estimated power consumption in watt/hour for different brands at the average line voltage. Replacing intensity (I) by current (C) in equation (5.4B), a pairwise comparison of different brands of same power is given in Table 16 using the linear model for current flow at the average line voltage.

Table 16: Pair-wise Comparison of Current flow of Different Brands by Power

Power	Brands	Difference in current flow	S.E. of the difference	d.f.	Calculated t-value
60	Phillips-Belco	0.0142	1.5069-04	8	94.2321***
	Phillips-Bengal	0.0160	1.6727-04	8	95.6515***
	Belco-Bengal	0.0018	1.5840-04	8	11.3635***
100	Phillips-Belco	0.0183	1.7006-04	8	107.6029***
	Phillips-Bengal	0.0317	3.1479-04	8	100.7001***
	Belco-Bengal	0.0134	2.8300-04	8	47.3482***

\*\*\* Difference is highly significant.

Analysis of Table 16 reflects that Phillips passes highest current followed by Belco and Bengal for both the power categories. In other words; brands can be ordered as per current flow as Phillips > Belco > Bengal at the average line voltage. Since average line voltage is constant at  $V = 215$ , it reflects that on the average, consumption of power is highest for Phillips and lowest for Bengal.

#### 5.10 Cost-Benefit Factors

Let  $T_{ij}$  be the average life of an electric bulb belonging to the  $j$ th brand and  $i$ th power category and  $x_{ij}$  be its price so that  $x_{ij}/T_{ij}$  be the cost of the bulb for unit time. Again let  $P_{ij}$  be the amount of power

consuming by the bulb in unit time at the average line voltage and  $Y_{ij}$  be its cost so that  $Z_{ij} = Y_{ij} + x_{ij}/T_{ij}$  be the total cost in unit time for burning the bulb. Now if  $L_{ij}$  be the amount of light emitting by the bulb in unit time at the average line voltage then  $Z_{ij}/L_{ij}$  be the cost per unit of light. In estimating the cost benefit factors, prices of bulbs are taken from the local retail market. Estimated average life  $T_{ij}$  are available from Table 4 of Chapter 4.  $P_{ij}$ 's are estimated from Table 15 of Chapter 5 at the average line voltage  $V = 215$  volt using the equation (5.5). Cost of power  $Y_{ij}$ 's are estimated as per current tariff of PDB (Power Development Board).  $L_{ij}$ 's are estimated by adjusting time unit from Table 10 of Chapter 5. Linear estimates of  $P_{ij}$ 's and  $L_{ij}$ 's are considered in estimating cost benefit factors. Table 17 shows the estimates of cost benefit factors.



Table 17: Estimated Cost-Benefit Factors at the Average Line Voltage  $V = 215$

power i	Brand j	Average life in hours $T_{ij}$	Average power consumption per month in KWH $P_{ij}$	Unit cost in Taka $x_{ij}$	cost of power per month in Taka $Y_{ij}$	Total cost per month in Taka $Z_{ij} = Y_{ij} + x_{ij}/T_{ij}$	Total amount of light av- ailable per month in (1.00000) $L_{ij}$	Cost per unit of light in Taka $Z_{ij}/L_{ij}$
60	Phillips	1925.6724	35.9600	16.00	79.80	85.79	401.6007	0.2136
	Belco	3270.1146	33.7619	13.00	75.77	78.63	304.7859	0.2579
	Bengal	1482.2653	33.4832	13.00	75.26	81.57	364.7493	0.2236
100	Phillips	2220.7779	59.8766	19.00	122.78	128.97	540.3665	0.2387
	Belco	2120.4358	57.0438	15.00	117.67	122.76	507.7866	0.2417
	Bengal	3525.4583	54.9695	15.00	113.95	117.01	545.0261	0.2147

## CHAPTER SIX

### SUMMARISATION OF RESULTS AND CONCLUDING REMARKS

#### 6.1 Summarisation of Results

Empirical survival probabilities for all categories of bulbs under investigation are given in Appendices IIA-IIE. Graphs displayed in Figures 2-7 (pp. 81-86 ) based on the survival probabilities that the underlying distributions of failure times for all the six categories of incandescent electric bulbs under investigation for tentatively follow two parameter negative exponential distribution. Results displayed in Table one (p. 87 ) also support that the underlying failure time distributions are two-parameter negative exponential, least square estimates of parameters are shown in Table two (p. 89 ). These estimates are biased and displayed for comparison only. Unbiased maximum likelihood estimates of parameters are portrayed in Table three (p. 95). A comparison of least square estimates and unbiased maximum likelihood estimates shows that in general least square over estimates the scale parameter (failure rate parameter) while it under estimates the location parameter (guarantee time). It is evident from Table three that of the 60 watt power category, Phillips has the highest guarantee time followed by Bengal and Belco. Their failure rates also follow the same order. In comparison to guarantee times and failure rates of other brands, the failure rate of Belco is observed to be much smaller. Of the 100 watt category Belco has the highest guarantee period followed by Phillips and Bengal. It is to be noted that for unbiased estimate, the guarantee time of

Bengal 100 watt is found to be statistically insignificant which is a contradiction with graphical result (pp. 81-86 ) and the result shown in Table one (p. 87 ). However, for uniformity, this insignificant result has been considered in further analysis. In this power category also, the failure rates follow the same ordering as the guarantee time. The unbiased estimates of mean life times for all the brand-categories under study are given in Table four including a 95% asymptotic confidence interval. The results are self explanatory. Of the 60 watt category, Belco has the highest estimated average life followed by Phillips and Bengal while Phillips has the minimum coefficient of variation followed by Bengal and Belco. Of the 100 watt category, Bengal has the highest estimated life time followed by Phillips and Belco while Belco has the minimum coefficient of variation followed by Phillips and Bengal. Pairwise comparison of estimated mean life are shown in Table five (p. 98 ) which reflects that the larger estimated average life for Belco 60 watt over Phillips and Bengal is statistically significance while Phillips and Bengal do not differ significantly for 100 watt, brands do not differ among themselves with respect to estimated average life. The likelihood ratio statistic for goodness of fit of fitted distributions using both unbiased maximum likelihood estimates and the least-square estimates are given in Table six (p. 101 ). The results reveal that the data are well graduated for all the six categories with both least square and unbiased m.l. estimates. Results of Mantel-Haenzel (1959) test for pair-wise comparison of survival curves of brands within the power category are displayed in Table seven (p. 104). The results are in agreement with the results of Table five (p. 98 ), that is, for 60 watt, the survival curve of Belco differs significantly

from Phillips and Bengal while Phillips and Bengal do not differ significantly and for 100 watt, the brands do not differ with one another significantly.

Relationship between voltage of electric line and the intensity of light has been discussed in section (5.3) and found as -

$$E (I/60 \leq V \leq 240) = b_0 + b_1V + b_2V^2 + b_3V^3$$

where I is the amount of light emitting per second, V is the line voltage,  $b_0$ ,  $b_1$ ,  $b_2$  and  $b_3$  are constant. No light was available for line voltage below 60 and no reading on light could be taken beyond line voltage 240 due to limitations of instruments. Fitted polynomials with standard error of estimates of  $b_0$ ,  $b_1$ ,  $b_2$  and  $b_3$  are shown in Table eight (p. 116). Practically line voltage varies in the range of 190-240 volts in general. In this range of line voltage, the voltage intensity relationship is found as

$$E (I/190 \leq V \leq 240) = a + bV$$

Fitted linear relationship for different brands and power along with the standard error of estimates are displayed in Table nine (p. 117). The intercept term for all the brands and power are found to be negative which is consistent with the fact that a certain amount of power is lost for heating the filament before emitting light. The scale parameter b is the rate of increase of intensity for unit change in the line voltage. For 60 watt power category, the rate of increase of intensity of light is found to be highest for Bengal followed by Phillips and Belco while

for 100 watt, it is found to be highest for Belco followed by Bengal and Phillips apparently.

The apparent wastage of electricity is found to be highest for Bengal followed by Belco and Phillips in order for 60 watt, while it is highest for Belco followed by Bengal and Phillips for 100 watt. The ratio  $(- a/b)$  may be considered as an index for wastage of electricity which is found to be highest for Belco followed by Bengal and Phillips for both 60 watt and 100 watt.

Estimated intensity of light at the average line voltage are displayed for both linear and polynomial relationships in Table ten (p. 118) and found to differ insignificantly. For this reason, the linear relationship is considered for further analysis. Pairwise comparison of parameters for different brands of same power are displayed in Table eleven (p. 121). For 60 watt, magnitude of intercept parameter is significantly larger than those of Phillips and Belco, while Phillips and Belco do not differ significantly. Side by side, the scale parameters of Belco is found to be significantly smaller than those of Phillips and Bengal, while Phillips and Bengal do not differ significantly. The Brands do not differ from one another significantly with respect to parameters in case of 100 watt.

A comparison of intensity of light at the average line voltage for different brands by power category is given in Table twelve (p. 122). It is observed that for 60 watt power category all the brands differ significantly with each other and their relative ordering is Phillips > Bengal > Belco. For 100 watt category, Belco has significantly smaller

intensity than those of Phillips and Bengal, while Phillips and Bengal do not differ significantly.

Voltage-current relationship has been studied in Section 5.7 and found to be a polynomial of third degree, that is,

$$E(C/0 < V < 240) = \beta_1 V + \beta_2 V^2 + \beta_3 V^3$$

where  $C$  is the amount of current flowing through the filament of the bulb for a given line voltage  $V$  in the range of 0-240 volts. Fitted polynomials are displayed in Table thirteen (p. 127). It is mentioned earlier that in reality, line voltage varies in the range of 190-240 volts generally. Current-voltage relationship in the range of 190-240 volts is found to be a linear one, that is,

$$E(C/190 \leq V \leq 240) = \alpha + \beta V$$

Fitted linear relationships for different brands and power categories of incandescent electric bulbs are appended in Table fourteen (p. 128). Estimated flow of current at the average line voltage as per polynomial relationship and linear relationship are shown in Table fifteen (p. 129) from where it is evident that the polynomial under estimates the observed flow of current at the average line voltage. Again, it over estimates the flow of current at the upper end points of line voltage. These facts instigated us to consider the linear relationship for further investigation. This table also reflects that flow of current is highest for Phillips and lowest for Belco for both 60 watt and 100 watt power categories of bulbs apparently. Table sixteen (p. 131) confirms these result, where pair-wise comparison has been made by power categories.

All the estimated cost-benefit factors including cost per unit of light at the average line voltage are appended in Table seventeen (p. 133). This table indicates that cost per unit of light is highest for Belco for both 60 watt and 100 watt bulbs. In case of 60 watt bulbs, minimum cost goes to Phillips while for 100 watt bulbs minimum cost goes to Bengal. For 60 watt bulbs, relative costs of Belco and Bengal are 17.18 per cent and 13.30 per cent higher than Phillips, while the relative cost of Bengal is 4.47 per cent higher than the Phillips. The relative cost of Belco and Phillips at the average line voltage in order are 11.17 per cent and 10.05 per cent higher in comparison to Bengal for 100 watt power category, while it is 1.24 per cent higher for Belco in comparison to Phillips.

## 6.2 Concluding Remarks

Apparently one should decide to purchase that bulb for which cost per unit light is minimum. But there are other factors to consider. Of these factors, first comes the guarantee period. Since average life of an electric bulb is not too small, consumers would expect a longer guarantee period. Second, cost of a bulb. If cost of a bulb is significantly higher than an alternative brand, consumers should prefer the low cost one. Above all, the room space where bulbs are used for light is an important factor which we did not consider at all. The optimality of light with respect to room space may convert a low cost bulb into high cost one. Some one may study this factor for further investigation. Considering the factors discussed above, it is not possible to say

explicitly which one is to prefer but preference will vary from consumer to consumer depending on their necessity, availability of alternatives, ability to invest for a bulb and knowledge of guarantee period, average life and the intensity of light. Phillips yields its maximum capacity at 250 volts. We have calculated the cost per unit of light at the average line voltage  $V = 215$ . If this average line voltage is changed, there will be a change in cost per unit light and that change may not be proportional to change in voltage which may upset the present results. In establishing the voltage-intensity relationship in Section 5.3 and voltage-current relationship in Section 5.7, we have considered the average of 3 observations on intensity of light and flow of current for the same line voltage. Instead, if exact observations could be used some difference may be expected in the estimated parameters which may effect the cost per unit of light. These are left as further scope for future investigation.



## APPENDIX - I

Consumption Pattern of Incandiscent Electric Bulbs  
(% of sell volume)

Sl. No.	Brands	Power	
		60 watt	100 watt
1.	Belco	16	17
2.	Bengal	9	10
3.	Crown	6	6
4.	Emon	-	-
5.	NEC	15	14
6.	Paramount	7	6
7.	Phillips	20	19
8.	Seiko	5	5
9.	Singer	14	14
10.	Sainik	8	9

- Records no sell during the short survey period of 3 days  
in the selected retailer shops.

## APPENDIX II

LOSS TIME ACCOUNT

Date	Time of loss		Amount of loss time in hour	Cumulative sum.
	From	To		
1-4-89	3.15 p.m.	5.45 p.m.	2.50	2.50
5-4	7.05 a.m.	8.20 a.m.	1.25	3.75
6-4	9.00 p.m.	10.45 p.m.	1.75	5.50
8-4	11.30 a.m.	2.30 p.m.	3.00	8.50
8-4 to 10-4	7.15 p.m.	1.15 p.m.	42.00	50.50
29-4	7.00 p.m.	8.30 p.m.	1.50	52.00
30-4	9.00 a.m.	10.00 a.m.	1.00	53.00
3-5	11.00 a.m.	12.00 a.m.	1.00	54.00
4-5	7.30 a.m.	8.15 a.m.	0.45	54.75
7-5	11.00 a.m.	12.15 a.m.	1.15	56.00
9-5	8.30 p.m.	10.00 p.m.	1.30	57.50
11-5	5.15 p.m.	7.15 p.m.	2.00	59.50
12-5	7.00 p.m.	9.30 p.m.	2.30	62.00
14-5	4.15 p.m.	5.15 p.m.	1.00	63.00
16-5	10.30 a.m.	1.30 p.m.	3.00	66.00
18-5	7.30 p.m.	10.00 p.m.	2.30	68.50
19-5	9.45 a.m.	11.15 a.m.	1.30	70.00
22-5	7.00 p.m.	10.30 p.m.	3.30	73.50
31-5	9.45 a.m.	12.45 a.m.	3.00	76.50
1-6-89	5.00 a.m.	7.00 a.m.	2.00	78.50
2-6			3.50	82.00
3-6-89	3.30 p.m.	5.30 p.m.	3.00	85.00
4-6	11.00 a.m.	12.45 a.m.	1.45	86.75
6-6	9.55 p.m.	9.30 p.m.	0.15	87.00

(contd.)

LOSS TIME ACCOUNT (contd.)

10-6	3.15 p.m.	6.00 p.m.	2.45	89.75
12-6	8.30 a.m.	10.00 a.m.	1.30	91.25
13-6	3.45 p.m.	7.30 p.m.	3.45	95.00
16-6	5.00 a.m.	5.45 a.m.	0.45	95.75
17-6	11.30 a.m.	1.30 p.m.	2.00	97.75
18-6	9.15 p.m.	10.30 p.m.	1.15	99.00
21-6	11.00 a.m.	1.00 p.m.	2.00	101.00
24-6	9.15 p.m.	11.15 p.m.	2.00	103.00
25-6	9.00 a.m.	11.30 a.m.	2.30	105.50
27-6	1.00 p.m.	1.30 p.m.	0.30	106.00
2-7-89	11.00 a.m.	1.00 p.m.	2.00	108.00

---

## APPENDIX III

OBSERVED FAILURE TIMES & LIFE TIMES IN HOURSTable Appendix IIIA showing life times of Phillips-60W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time within observation period	Exact life time in hour
34	3-6-89	7.00 a.m.	1516	82	1434
116	"	"	"	"	"
40	4-6-89	3.00 p.m.	1548	85	1463
79	6-6-89	7.00 a.m.	1588	87	1501
49	7-6-89	3.00 p.m.	1620	87	1533
70	9-6-89	7.00 a.m.	1660	87	1573
01	13-6-89	3.00 p.m.	1764	91.25	1672.75
82	"	"	"	"	"
55	16-6-89	7.00 a.m.	1828	95.75	1732.25
73	"	"	"	"	"
52	19-6-89	3.00 p.m.	1908	99	1809
85	22-6-89	7.00 a.m.	1972	1.1	1871
43	28-6-89	3.00 p.m.	2124	106	2018
76	"	"	"	"	"
04	2-7-89	11.00 p.m.	2228	108.25	2119.75

## APPENDIX IIIB

Table showing life times of Belco-60W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time within observation period	Exact life time in hour
119	28-4-89	3.00 p.m.	660	50.50	609.50
57	6-5-89	3.00 p.m.	852	54.75	797.25
120	12-5-89	3.00 p.m.	996	59.50	936.50
33	20-5-89	3.00 p.m.	1188	70	1118
87	30-5-89	11.00 p.m.	1436	73.50	1362.50
09	6-6-89	11.00 p.m.	1604	87	1517
63	17-6-89	7.00 a.m.	1852	95.75	1756.25
39	26-6-89	11.00 p.m.	2084.10	105.50	1978.50
10	4-7-89	3.00 p.m.	2268	108	2160

## APPENDIX IIIC

Table showing life times of Bengal-60W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time within observation period	Exact life time in hour
110	30-4-89	3.00 p.m.	708	53	655
58	"	"	"	"	"
80	4-5-89	11.00 p.m.	812	54.75	757.25
47	"	"	"	"	"
62	715	11.00 p.m.	884	56	828
05	10-5-89	7.00 a.m.	940	57.50	882.50
74	12-5-89	11.00 p.m.	1004	62	962
48	15-5-89	3.00 p.m.	1068	63	1035
71	20-5-89	3.00 p.m.	1188	70	1118
08	23-5-89	11.00 p.m.	1268	73.50	1194.50
77	28-5-89	7.00 a.m.	1372	73.50	1298.50
38	1-6-89	7.00 a.m.	1468	78.50	1389.50
89	5-6-89	7.00 a.m.	1564	86.75	1477.25
33	12-6-89	11.00 p.m.	1748	91.25	1656.75
101	20-6-89	11.00 p.m.	1940	99	1841
35	28-6-89	3.00 p.m.	2124	106	2018

## APPENDIX IIID

Table showing life times of Phillips-100W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time observation period	Exact life time in hour
07	9-6-89	3.00 p.m.	1668	87	1581
16	11-6-89	7.00 a.m.	1708	89.75	1618.25
46	13-6-89	11.00 p.m.	1772	95	1677
61	"	"	"	"	"
118	17-6-89	3.00 p.m.	1860	97.75	1762.25
19	21-6-89	11.00 p.m.	1964	101	1863
28	"	"	"	"	"
103	23-6-89	3.00 p.m.	2004	101	1903
22	26-6-89	7.00 p.m.	2068	105.50	1962.50
94	28-6-89	7.00 a.m.	2116	106	2010
37	1-7-89	7.00 a.m.	2188	106	2082
25	4-7-89	7.00 a.m.	2260	108.25	2151.75

## APPENDIX III E

Table showing life times of Belco-100W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time within observation period	Exact life time in hour
23	11-6-89	3.00 p.m.	1716	89.75	1626.25
11	"	"	"	"	"
17	"	"	"	"	"
98	13-6-89	7.00 a.m.	1756	91.25	1664.75
59	14-6-89	3.00 p.m.	1788	95	1693
02	17-6-89	11.00 p.m.	1868	97.75	1770.25
53	19-6-89	7.00 a.m.	1900	99	1801
14	20-6-89	11.00 p.m.	1940	99	1841
107	23-6-89	11.00 p.m.	2012	101	1911
56	26-6-89	7.00 a.m.	2068	105.50	1962.50
50	28-6-89	3.00 p.m.	2124	106	2018
41	2-7-89	7.00 a.m.	2212	106	2106



## APPENDIX III F

Table showing life times of Bengal-100W

Sample No.	Date of failure	Time of failure	Observed life time in hour	Loss time within obser- vation period	Exact life time in hour
51	11-4-89	3.00 p.m.	252	50.50	201.50
42	19-4-89	11.00 p.m.	252	50.50	401-50
99	28-4-89	3.00 p.m.	660	50.50	609.50
15	7-5-89	7.00 a.m.	868	54.75	813.25
60	17-5-89	11.00 p.m.	1124	66	1058
30	28-5-89	7.00 a.m.	1372	73.50	1298.50
12	9-6-89	11.00 p.m.	1676	87	1589
03	22-6-89	3.00 p.m.	1980	101	1879
18	4-7-89	3.00 p.m.	2268	108	2160

## APPENDIX IVA

Table of survival probability in case  
of Phillips-60W

Life time	Survival prob. obtained from Kaplan Meier estimate	Survival prob. when parameters are obtained from m.l.e	Survival prob. when parameter are obtained from regression
1430.00	0.90	.9586	.9116
1463.00	0.85	.8993	.8569
1501.00	0.80	.8356	.7978
1533.00	0.75	.7855	.7513
1573.00	0.70	.7271	.6970
1672.75	0.60	.5996	.5779
1732.25	0.50	.5344	.5171
1809.00	0.45	.4607	.4475
1871.00	0.40	.4087	.3983
2018.00	0.30	.3076	.3022
2119.75	0.25	.2530	.2497

## APPENDIX IVB

Table of survival probability in case  
of Belco-60W

Life time	Survival prob. obtained from Kaplan Meier estimate	Survival prob. when parameters are obtained from m.l.e	Survival prob. when parameter are obtained from regression
609.50	0.95	0.9512	0.9527
797.25	0.90	0.8895	0.8954
936.50	0.85	0.8464	0.8504
1118.00	0.80	0.7932	0.7986
1362.50	0.75	0.7269	0.7382
1517.00	0.70	0.6879	0.6957
1756.25	0.65	0.6316	0.6418
1978.50	0.60	0.5834	0.5932
2160.00	0.55	0.5468	0.5522

## APPENDIX IVC

Table of survival probability in case  
of Bengal-60W

Lite time in hour	Survival prob. obtained from Kaplan Meier estimate	Survival prob. when parameters are obtained from m.l.e.	Survival prob. when parameter are obtained from regression
655.00	0.90	0.9512	0.9039
757.25	0.80	0.8458	0.8075
828.00	0.75	0.7798	0.7469
882.50	0.70	0.7325	0.7034
962.00	0.65	0.6686	0.6444
1035.00	0.60	0.6148	0.5945
1118.00	0.55	0.5589	0.5426
1194.50	0.50	0.5119	0.4987
1298.50	0.45	0.4543	0.4447
1389.50	0.40	0.4092	0.4022
1477.25	0.35	0.3700	0.3652
1656.75	0.30	0.3010	0.2996
1841.00	0.25	0.2436	0.2445
2018.00	0.20	0.1988	0.2012

## APPENDIX IV D

Table of survival probability in case of Phillips-100w

Life time in hour	Survival prob. when parameters are obtained from Kaplan- Meier estimate	Survival prob. when parameters are obtained from m.l.e	Survival prob. when parameter are obtained from regression
1588.00	0.95	0.9413	0.9468
1618.25	0.90	0.9000	0.9051
1677.00	0.80	0.8248	0.8297
1762.25	0.75	0.7267	0.7303
1863.00	0.65	0.6257	0.6285
1903.00	0.60	0.5897	0.5921
1962.50	0.55	0.5398	0.5419
2010.00	0.50	0.5030	0.5049
2082.00	0.45	0.4520	0.4535
2151.75	0.40	0.4075	0.4087

## APPENDIX IVE

Table of survival probability in case of Belco-100W

Life time in hour	Survival prob. when parameters are from Kaplan-Meier estimate	Survival prob. when parameters are obtained from m.l.e	Survival prob. when parameters are obtained from regression
1626.25	0.85	0.9532	0.8543
1664.75	0.80	0.8850	0.8034
1693.00	0.75	0.8381	0.7680
1770.25	0.70	0.7222	0.6789
1801.00	0.65	0.6807	0.6464
1841.00	0.60	0.6302	0.6064
1911.00	0.55	0.5507	0.5423
1962.50	0.50	0.4987	0.4995
2018.00	0.45	0.4481	0.4571
2106.00	0.40	0.3782	0.3972

## APPENDIX IVF

Table of survival probability in case of Bengal-100W

Life time in hour	Survival prob. when parameters are obtained from Kaplan-Meier estimate	Survival prob. when parameters are obtained from m.l.e	Survival prob. when parameters are obtained from regression
201.50	0.95	0.9512	0.9505
401.50	0.90	0.8983	0.8993
609.50	0.85	0.8465	0.8490
813.75	0.80	0.7985	0.8023
1058.00	0.75	0.7446	0.7498
1298.50	0.70	0.6953	0.7015
1589.00	0.65	0.6398	0.6473
1879.00	0.60	0.5889	0.5973
2160.00	0.55	0.5434	0.5526

## APPENDIX - V

Average Intensity of Light for Given Line Voltage  
in Lux per Second

Given voltage	Average Intensity					
	60 watt bulbs			100 watt bulbs		
	Phillips	Belco	Bengal	Phillips	Belco	Bengal
60	0.50	0.00	0.80	1.40	1.10	1.60
70	1.30	0.37	1.63	2.90	2.20	2.80
80	2.60	1.13	2.27	5.50	4.40	5.30
90	4.80	2.27	3.83	10.07	7.63	9.47
100	8.20	4.50	6.60	16.03	12.60	15.60
110	13.23	7.73	10.43	25.07	20.10	23.97
120	19.97	12.37	15.87	36.27	29.47	35.37
130	28.23	18.40	22.57	43.10	42.47	49.64
140	40.97	24.87	31.17	70.70	57.60	68.80
150	52.30	33.77	41.57	88.50	76.60	88.10
160	68.30	44.27	58.07	110.00	96.10	109.50
170	86.00	57.50	74.60	131.50	116.70	131.90
180	104.00	71.27	90.50	153.40	130.00	153.80
190	121.40	86.63	104.50	174.30	158.20	174.00
200	139.70	101.63	124.87	194.40	176.30	194.70
210	158.30	118.50	143.77	213.80	197.90	214.70
220	175.90	135.25	161.00	235.60	216.80	234.30
230	193.10	149.37	178.50	253.50	236.80	255.30
240	211.20	167.27	195.27	273.40	257.90	283.60



## APPENDIX - VI

Average Electricity Passing Through the Filament for  
Given Line Voltage in Ampere Per Hour

Given Voltage	Average Electricity Flow					
	60 Watt bulbs			100 Watt bulbs		
	Phillips	Belco	Bengal	Phillips	Belco	Bengal
0	0	0	0	0	0	0
1	0.0033	0.0027	0.0030	0.0053	0.0047	0.0047
3	0.0083	0.0080	0.0083	0.0153	0.0137	0.0140
5	0.0133	0.0130	0.0133	0.0250	0.0223	0.0227
7	0.0183	0.0177	0.0183	0.0347	0.0310	0.0317
10	0.0257	0.0250	0.0257	0.0483	0.0433	0.0443
15	0.0377	0.0360	0.0373	0.0703	0.0633	0.0643
20	0.0483	0.0467	0.0487	0.0910	0.0817	0.0830
30	0.687	0.0660	0.0683	0.1283	0.1153	0.1177
40	0.0860	0.0827	0.0857	0.1607	0.1447	0.1473
50	0.1017	0.0973	0.1003	0.1883	0.1700	0.1733
60	0.1150	0.1100	0.1130	0.2027	0.1920	0.1927
70	0.1250	0.1200	0.1227	0.2207	0.2080	0.2100
80	0.1347	0.1280	0.1310	0.2373	0.2240	0.2260
90	0.1440	0.1360	0.1387	0.2503	0.2390	0.2400
100	0.1527	0.1440	0.1460	0.2663	0.2530	0.2537
110	0.1613	0.1517	0.1530	0.2793	0.2667	0.2667
120	0.1693	0.1590	0.1600	0.2913	0.2787	0.2790
130	0.1767	0.1660	0.1667	0.3023	0.2893	0.2900
140	0.1840	0.1730	0.1730	0.3127	0.2993	0.3000
150	0.1913	0.1800	0.1797	0.3233	0.3090	0.3090
160	0.1980	0.1857	0.1857	0.3337	0.3190	0.3170
170	0.2053	0.1917	0.1927	0.3430	0.3287	0.3247
180	0.2117	0.1977	0.1970	0.3537	0.3373	0.3310
190	0.2177	0.2033	0.2027	0.3630	0.3467	0.3380
200	0.2237	0.2097	0.2080	0.3727	0.3550	0.3440
210	0.2293	0.2153	0.2133	0.3820	0.3640	0.3523
220	0.2357	0.2210	0.2190	0.3910	0.3730	0.3580
230	0.2413	0.2267	0.2247	0.4017	0.3820	0.3650
240	0.2467	0.2327	0.2307	0.4107	0.3910	0.3737

REFERENCES

## REFERENCES

- BAIN, L.J. ENGELHARDT, M. and WRIGHT, F.T. (1985). Test for increasing trend in the intensity of a poisson Process; A power study, JASA, vol. 80, No. 390, pp. 419-422.
- BARLOW, R.E. BARTHOLOMEW, D.J., BREUNNER, J.M. and BRUNK, H.D. (1972). Statistical inference under order Restrictions, John Wiley, New York.
- BUCKLAND, W.R. (1964). Statistical Assessment of the life characteristics. Griffin, London.
- COX, D.R. (1978). Some remarks on the role in statistics of graphical methods. Applied Statistics, 27, pp. 4-9.
- CRAMMER, H. (1946). Mathematical Methods of Statistics, Princeton, New Jersey.
- CROW, L.H. (1974). Reliability analysis for complex, repairable systems. In Reliability and Biometry, SIAM.
- CHEN, Y.Y., HOLLANDER, M. and LANGBERG, N.A. (1983). Testing whether new is better than used with randomly censored data, Annals of statistics, V. II, pp. 267-74.
- DAVIS, D.J. (1952). An analysis of some failure data. JASA, Vol. 47, pp. 113-50.
- EPSTEIN, B. (1958). The exponential distribution and its role in life testing. Industrial Quality Control, Vol. 15, pp. 2-7.
- EPSTEIN, B. and M. SOBEL (1953). Life testing, JASA, Vol. 48, pp. 486-502.

- FEIGL, P. and M. ZELEN (1965). Estimation of exponential survival probabilities with Concomitant informations, *Biometrics*, Vol. 21, pp. 826-38.
- GAIL, M.H. and GASTWIRTH, J.L. (1978). A scale free Goodness of fit test for the exponential distribution based on the Gini Statistic. *JRSS, B* 40, pp. 350-357.
- GUMBIL, E.J. (1958). *Statistics of Extremes*, New York, Col. U.P.
- HEGAZY, Y.A.S. and GREEN, J.R. (1975). Some new goodness of fit tests using order statistics. *Appl. Stat.* 24, pp. 299-308.
- HOERL, A.E., KENNARD, R.W. and K.F. BALDWIN (1975). Ridge regression: Some simulations. *Communication in Statistics*, 4A, pp. 105-23.
- HOLLANDER, M. and PROSCHAN, F. (1979). Testing to determine the underlying distribution using randomly censored data. *Biometrics* 35, pp. 393-401.
- JOHNSON, N.L. and S. KOTZ (1970). *Continuous univariate distributions*. Boston, Massachusetts: Houghton Mifflin.
- KALBFLEISCH, J.D. and R.L. PRINTICE (1980). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons.
- KAPLAN, E.L. and P. MEIER (1958). Non parametric estimation from incomplete observations. *JASA*, 53, 457-81.
- KENDALL, M.G. and STAJART (1961). *Advanced theory of Statistics*, Vol. 2. Griffin, London.
- LAWLESS, J.F. (1982). *Statistical Models and Methods for life time data*. John Wiley and Sons.

- LITTEL, R.C., J. J. McCLAVE and W.W. OFTEN (1979). Goodness of fit tests for two parameter Weibull distribution, *Comm. Stat.*, B8, pp. 257-269.
- MANDENHALL, W. and HADER, R.J. (1958). Estimation of Parameters of a mixed exponentially distributed failure time distributions from Censored life test data. *Biometrika*, 45, 504-20.
- MANN, N.R. and FERTIG, K.W. (1975). A goodness of fit test for the two parameter Vs. Three parameter Weibull Confidence bound for the threshold. *Technometrics* 17, pp. 237-245.
- MANTEL, N. and W. HAENSZEL (1959). Statistical aspects of the analysis of data from retrospective studies of discuse. *J. Nat. Cancer Institute*, 22, 719-48.
- MICHAEL, J.R. and SCHUCANY, W.R. (1979). A new approach to testing goodness of fit with censored samples. *Technometrics*, 21, pp. 435-442.
- MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. (1974). Introduction to the theory of statistics, 3rd ed., McGraw-Hill, London.
- NELSON, W.B. and G.J. HAHN (1972). Linear estimation of a regression relationship from censored data. Part I - simple methods and their applications. *Technometrics*, 15, 133-150.
- NELSON, W.B. (1972). Theory and application of hazard plotting for Censored failure data. *Technometrics*, 14, 945-65.
- PETTIT, A.N. (1977). Test for the exponential distribution with censored data using cramer Von Mises statistics, *Biometrika*, 64, pp. 629-632.
- SHAPZRO, S.S., and WILK, M.B. (1965). An analysis of variance test for normality, *Biometrika*, 52, pp. 529-611.

SHOUMAN, M.L. (1968). Probabilistic Reliability: An Engineering Approach. New York: McGraw-Hill.

TIKU, M.L. and SINGH, M. (1981). Testing the two parameters weibull distribution, Comm. Stat. Series 10 A, pp.

TIKU, M.L. (1981). Goodness of fit statistic based on the spacings of complete or censored samples, Australian Journal of Statistics.

WEIBULL, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanich, 18, 293-297.

Rajshahi University Library  
Documentation Section  
Document No. ~~D-1584~~  
Date..23.2.92..... D-1584