A STUDY OF WEIBULL DISTRIBUTION

A Project report submitted in partial fulfillment of the requirements for the degree of M.Sc.(Statistics) with specialization in Industrial Statistics

Submitted by

Ms. Bhavsar Shweta Vijay (382753)

Ms. Patil Ashwini Diliprao (382780)

Mr. Hire Manish Nagraj (382762)

Project Guide

Dr. (Mrs.) Kirtee Kamalja

at the

Department of Statistics

School of Matehmatical Sciences

Kavayitri Bahinabai Chaudhari North Maharashtra University

Jalgaon

(June 2022)

CERTIFICATE

This is to certify that **Ms. Patil Ashwini Diliprao, Ms. Bhavsar Shweta Vijay and Mr. Hire Manish Nagraj** are the student of **M.Sc. Statistics** with specialization in Industrial Statistics at the Department of Statistics, School of Mathematical Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. They have successfully completed their project work entitled **"A Study of Weibull Distribution"** based on research paper reviews as a part of M.Sc. (Statistics) program under my guidance and supervision during the academic year 2021-2022.

Date:

Place: Jalgaon

Dr.(Mrs.) Kirtee K. Kamalja

(Project Guide)
Department of Statistics
K.B.C.North Maharashtra University
Jalgaon

ACKNOWLEDGEMENT

We want to convey our sincere regards to **Prof. R. L. Shinde**, Head, Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for seeking us the desire permission for this project.

We take this opportunity in expressing our sincere gratitude to our project guide **Dr.(Mrs.) Kirtee K. Kamalja** for her valuable guidance, kind suggestions, co-operation and constant encouragement, which enabled us to take every forward step in our project.

It would be unfair to go without acknowledging our beloved friends at the campus of Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for their precious support to us.

Last but not the least, we are thankful to our parents for their moral support and blessings.

Ms. Patil Ashwini Diliprao (382780)

Ms. Bhavsar Shweta Vijay (382753)

Mr. Hire Manish Nagraj (382762)

Contents

Chapter 1

Introduction

Weibull models are used to describe various types of observed failures of components and phenomena. They are widely used in reliability and survival analysis. In addition to the traditional two-parameter and three parameter Weibull distributions in the reliability or statistics literature, many other Weibull-related distributions are available. The purpose of this chapter is to give a brief introduction to those models, with the emphasis on models that have the potential for further applications. After introducing the traditional Weibull distribution, some historical development and basic properties are presented. We also discuss estimation problems and hypothesis-testing issues, with the emphasis on graphical methods. Many extensions and generalizations of the basic Weibull distributions are then summarized. Various applications in the reliability context and some Weibull analysis software are also provided.

The Weibull distribution is a continuous probability distribution named after Swedish mathematician Waloddi Weibull. He originally proposed the distribution as a model for material breaking strength, but recognized the potential of the distribution in his 1951 paper A Statistical Distribution Function of Wide Applicability. Today, it's commonly used to assess product reliability, analyze life data and model failure times. The Weibull can also fit a wide range of data from many other fields, including: biology, economics, engineering sciences, and hydrology (Rinne, 2008).

The Weibull distribution was perhaps best summarized in the 1951 paper [1] appropriately titled: " A Statistical Distribution Function of Wide Applicability" published in the Journal of Applied Mechanics. Nearly 50 years have passed since this event, and time has proven professor Weibull correct. The Weibull has found use in many quarters across a wide portion of the engineering and scientific community. Its popularity, diversity of application and theoretical development continues to increase.

1.1 History of Weibull distribution

In probability theory and statistics, the Weibull distribution is one of the most important continuous probability distributions. It was, first introduced by ***Professor Waloddi Weibull*** in 1939 when he was studying the issue of structural strength and life data analysis, and was formally named after him later in 1951. He proposed the "chain" model to explain the structural strength. Based on the assumption that a structure is composed of several small components (n pieces) in series, we could consider the structure as being composed of an n-rings chain, the strength of which (or life) completely depends on the weakest ring's strength (or life). In his model, with the assumption that the strength of different rings are independent and identically distributed, finding the strength distribution of the chain become the problem of finding the distribution of the weakest ring.

Due to the result of research conducted by Gnedenko (1943), no matter what the original distribution of the variable is, the asymptotic distribution of the minimum could only be three different forms. The Weibull distribution is one of them.

Since Weibull distribution is established on the weakest link

model, which could sufficiently reflect the defect of material and the effects of stress concentration, it has been considered as appropriate model to describe strength of fiber material in practical application.

1.2 Applications of Weibull Distribution

Weibull distribution has been used as a model in diverse disciplines to study many different issues. Some of them are:

- Because of its flexible shape and ability to model a wide range of failure rates, the Weibull has been used successfully in many applications as a purely empirical model.
- The Weibull model can be derived theoretically as a form of Extreme Value Distribution, governing the time to occurrence of the "weakest link" of many competing failure processes. This may explain why it has been so successful in applications such as capacitor, ball bearing, relay and material strength failures.
- In electrical engineering to represent overvoltage occurring in an electrical system.
- In industrial engineering to represent manufacturing and delivery times.
- In weather forecasting and the wind power industry to describe wind speed distributions, as the natural distribution often matches the Weibull shape.
- In hydrology the Weibull distribution is applied to extreme events such as annual maximum one-day rainfalls and river discharges.
- In describing the size of particles generated by grinding, milling and crushing operations, the 2-Parameter Weibull distribution is used.

Chapter 2

Weibull Distribution

2.1 What is Weibull Distribution?

The Weibull distribution is a continuous probability distribution that can fit an extensive range of distribution shapes. There are two versions of this distribution. The three-parameter Weibull distribution, unsurprisingly, has three parameters, shape, scale, and threshold. When the threshold parameter is zero, it is known as the two-parameter Weibull distribution. In this chapter we are going to study the pdf and cdf of Weibull distribution and plotting respective graphs using R-software.

2.2 Probability Density Function of Weibull Distribution

If X has a two-parameter Weibull distribution, then $Y = X + c$ has a three-parameter Weibull distribution with the added location parameter μ. The probability density function of a Weibull random variable is:

$$f_X(x) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{\alpha}}; & \text{if } \alpha > 0, \quad \beta > 0, \quad x \geq \mu \\ 0; & \text{if } x < \mu \end{cases}$$

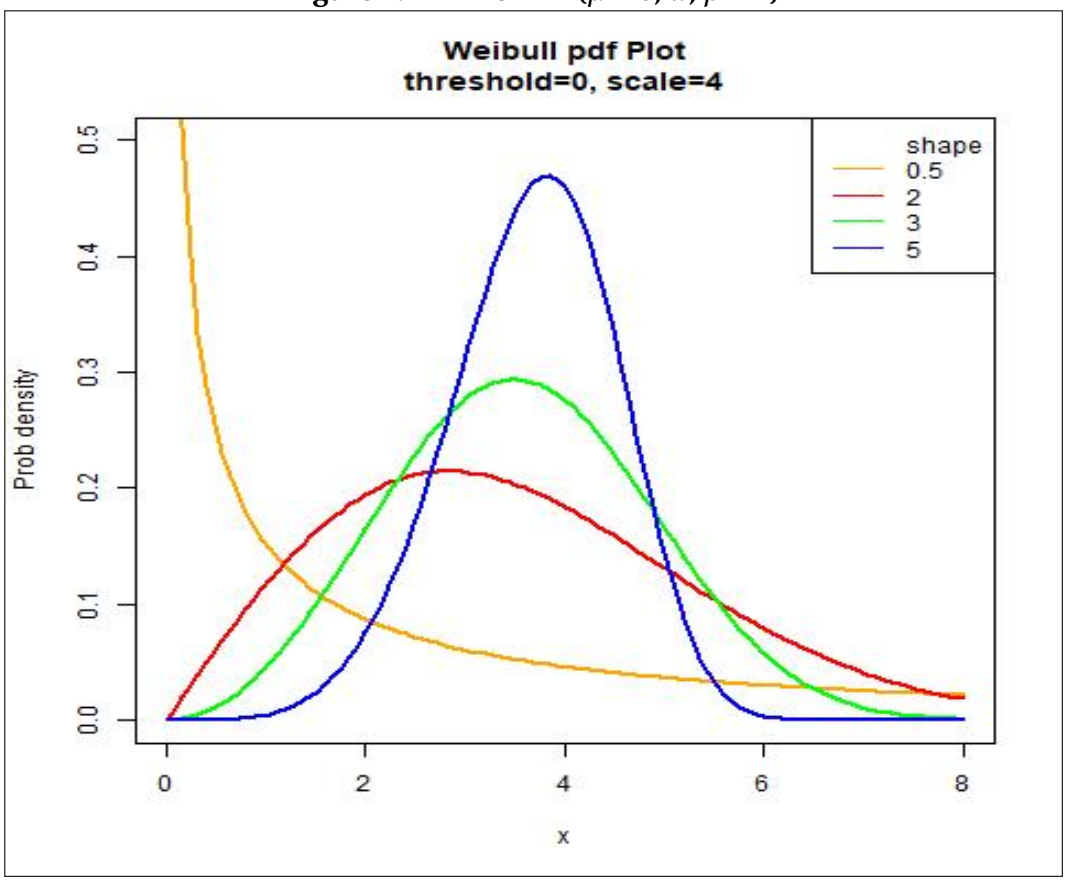
where, α is the shape parameter, β is the scale parameter and μ is the threshold or location parameter.

Notation: $X \sim Weibull(\mu, \alpha, \beta)$ or $X \sim WD(\mu, \alpha, \beta)$

2.2.1 Sketching pdf of WD($\mu = 0, \alpha, \beta = 4$)

Unsurprisingly, the parameter α describes the shape of the data's distribution. It is also refer as the Weibull slope because its value equals to the slope of the line on a probability plot. In these probability distribution plots, μ and β are constants to highlight the impact of changing the shape.

Figure 2.1 PDF of WD($\mu = 0$, α, $\beta = 4$)



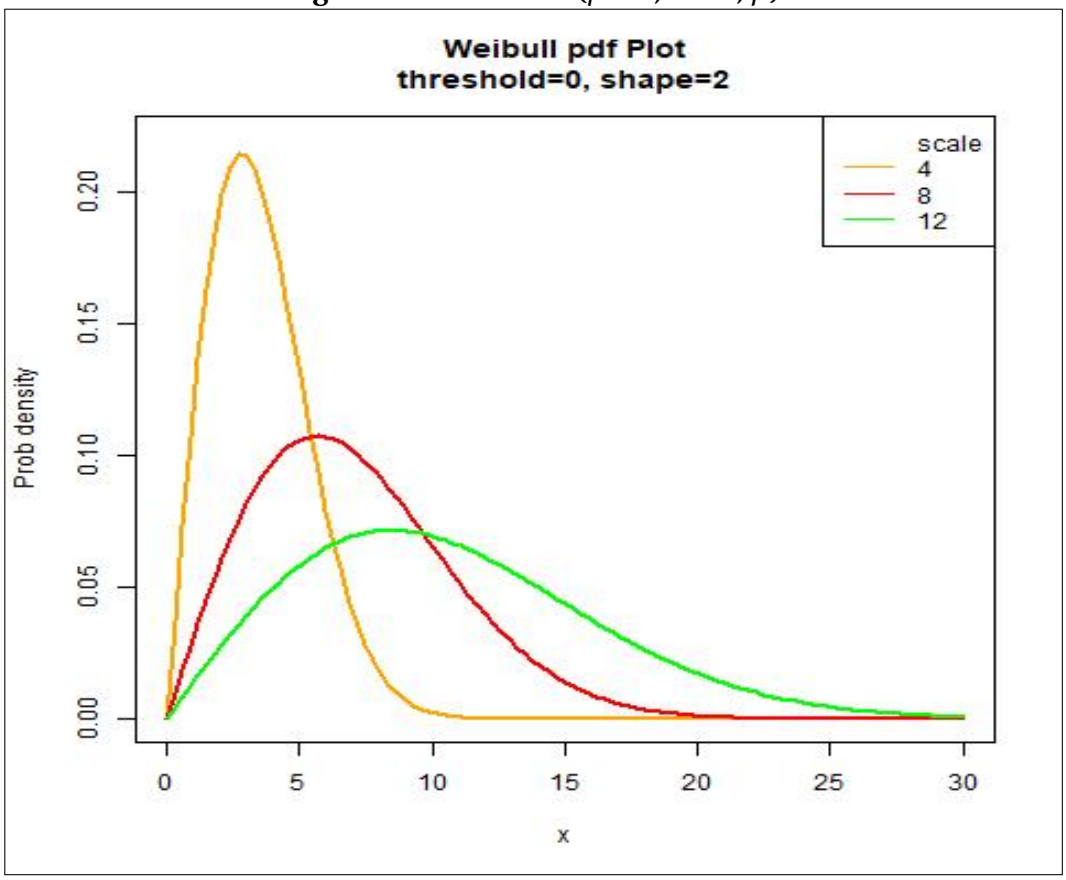
Observations: The following is the interpretation of above graph.

- When $\alpha < 1$ The Weibull distribution has steadily decreasing values.
- When $\alpha = 3$ it approximates a normal distribution.
- When $\alpha > 3.7$ it is Left-skewed.

2.2.2 Sketching pdf of WD($\mu = 0, \alpha = 2, \beta$)

The parameter β represents the variability present in the distribution. Changing the β parameter affects how far the probability distribution stretches out. As you increase the β, the distribution stretches further right, and the height decreases. Decreasing the scale shrinks the distribution to the left and increases its peak, as shown below in figure 2.2.

Figure 2.2 PDF of WD($\mu = 0$, $\alpha = 2$, β)



Observations: The value of the β parameter equals the 63.2 percentile in the distribution i.e. 63.2% of the values in the distribution are less than the β value.

2.3 CDF of Weibull Distribution

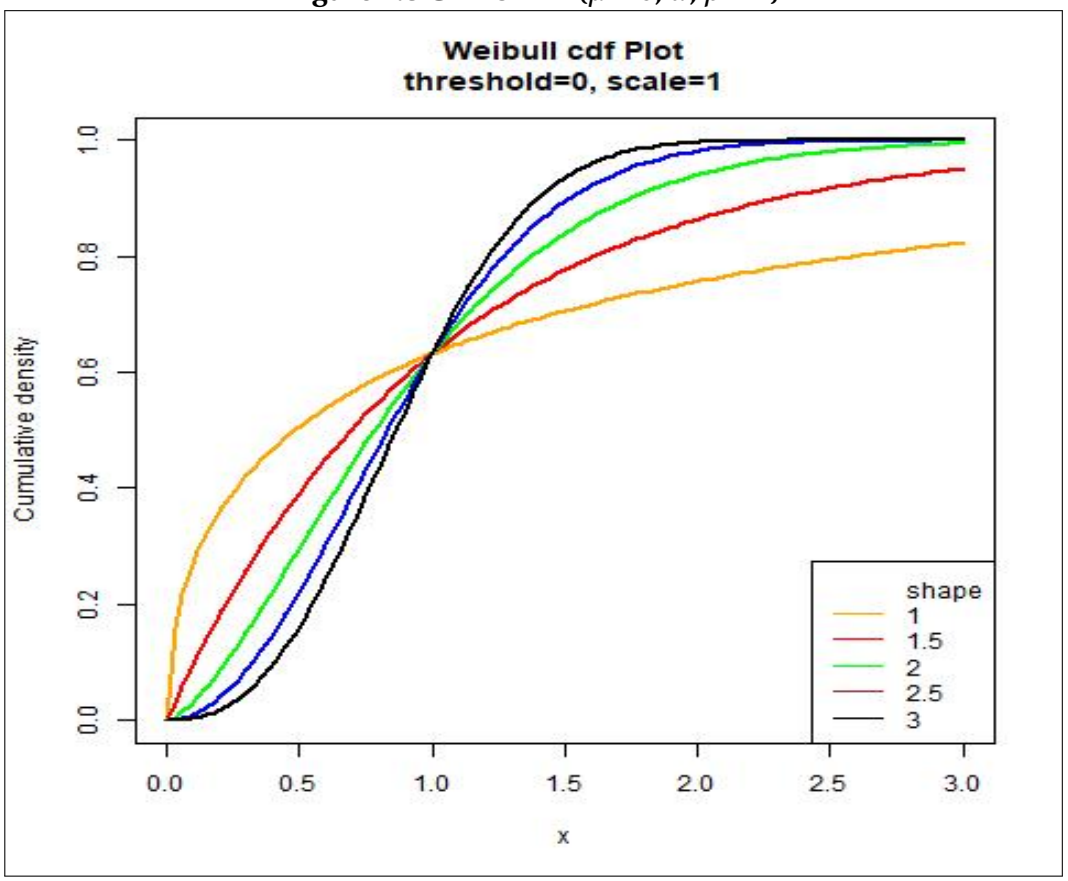
The cumulative distribution function for the Weibull distribution is:

$$F_X(x) = \begin{cases} 1 - e^{-\left(\frac{x-\mu}{\beta}\right)^{\alpha}}; & \text{if } \beta > 0, \quad x \geq \mu \\ 0; & \text{if } x < \mu \end{cases}$$

2.3.1 Sketching CDF of WD($\mu = 0, \alpha, \beta = 1$)

To visualize the CDF of Weibull distribution for various values of α by holding μ and β constant, we sketch these by using R Software. It is shown in the following Figure 2.3.

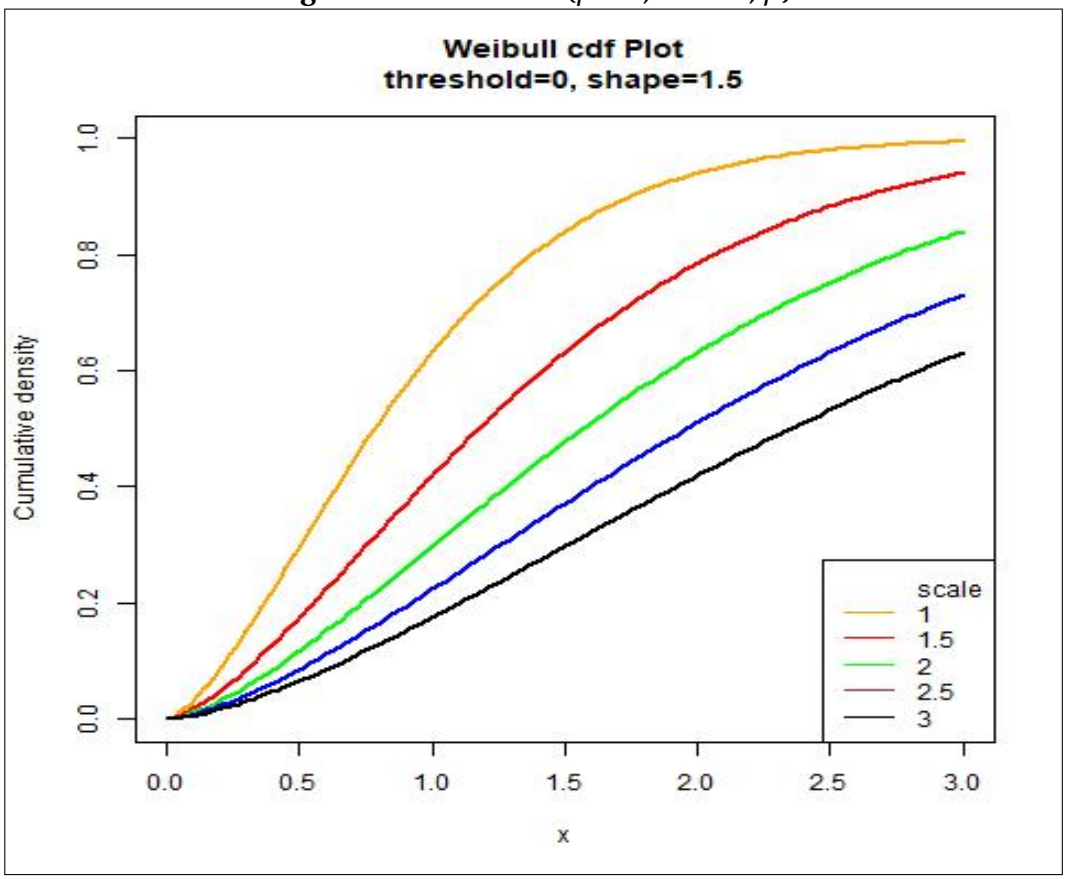
Figure 2.3 CDF of WD($\mu = 0$, α, $\beta = 1$)



2.3.2 Sketching CDF of WD($\mu = 0, \alpha = 1.5, \beta$)

To visualize the CDF of Weibull distribution for various values of β by holding μ and α constant, we sketch these by using R Software. It is shown in the following Figure 2.4.

Figure 2.4 CDF of WD($\mu = 0$, $\alpha = 1.5$, β)



2.4 Particular cases of Weibull Distribution

- When $X \sim Weibull(\mu = 0, \alpha, \beta = C)$ where C is an assumed constant value then it follows the 1-parameter Weibull distribution. Here the only unknown parameter is the scale parameter β. Note that in the formulation of the 1-parameter Weibull, we assume that the shape parameter α is known a priori from past experience with identical or similar products.

- When $X \sim Weibull(\mu = 0, \alpha, \beta)$ then it follows 2-parameter Weibull distribution. It is also known as standard Weibull model.
- When $X \sim Weibull(\mu = 0, \alpha, \beta = 1)$ then it is referred as the standard Weibull distribution.
- When $X \sim Weibull(\mu, \alpha = 1, \beta)$ then it follows the Exponential Distribution.
- When $X \sim Weibull(\mu, \alpha = 2, \beta)$ then it follows the Rayleigh Distribution.

2.5 Statistical Properties of Weibull Distriution

Various measures of WD(μ, α, β) are given in the following table.

Table 2.1

Population Quantity	**WD**$(\mu = 0, \alpha, \beta = 1)$	**WD**$(\mu = 0, \alpha, \beta)$	**WD**(μ, α, β)
$E(X)$	$\Gamma(1+\frac{1}{\alpha})$	$\beta\Gamma(1+\frac{1}{\alpha})$	$\mu+\beta\Gamma(1+\frac{1}{\alpha})$
Median	$ln\,(2)^{\frac{1}{\alpha}}$	$\beta ln\,(2)^{\frac{1}{\alpha}}$	$\mu+\beta ln\,(2)^{\frac{1}{\alpha}}$
Mode	$(1-\frac{1}{\alpha})^{\frac{1}{\alpha}};$ if $\alpha > 1$	$\beta(1-\frac{1}{\alpha})^{\frac{1}{\alpha}};$ if $\alpha > 1$	$\mu+\beta(1-\frac{1}{\alpha})^{\frac{1}{\alpha}};$ if $\alpha > 1$
σ	$\sqrt{\Gamma(\frac{\alpha+2}{\alpha}) - (\Gamma(\frac{\alpha+1}{\alpha}))^2}$	$\beta\sqrt{\Gamma(\frac{\alpha+2}{\alpha}) - (\Gamma(\frac{\alpha+1}{\alpha}))^2}$	$\beta^2\Gamma(1+\frac{2}{\alpha}) - \mu^2$

Chapter 3

R-Packages for Weibull Distribution

For the simulation study and fitting of the Weibull distribution, we use R-software. A brief summary of the commands which are the codes used for simulation and fitting of the Weibull Distribution are given in this chapter.

3.1 Comparision of pdf and CDF with respect to R-software

The pdf and cdf of a Weibull random variable X in R software is as same as pdf and CDF given in chapter 2 respectively. In R-software the location parameter is denoted by θ.

3.2 Packages in R-software for Weibull Distribution

In the R software, for the fitting of Weibull Distribution we are using the following packages which can be listed as follows:

- FAdist
- Weibullness
- FITDISTRPLUS
- Weibullfit
- ForestFit

Both the packages FITDISTRPLUS and Weibullness uses WD(α, β, θ). A brief summary of each package is given in the following sections.

3.2.1 FAdist

This package contains several distributions that are sometimes useful in hydrology. It helps to calculate density, distribution function, quantile function and random generation for the 3-parameter WD(μ, α, β). Following are the commands in the FAdist package:

Table 3.1 Commands used in FAdist package

Sr. No.	Command	Syntax	Purpose
1	dweibull3	dweibull3(x, shape, scale=1, thres=0, log=FALSE)	gives the density function of Weibull distribution
2	pweibull3	pweibull3(q, shape, scale=1, thres=0, lower.tail=TRUE, log.p=FALSE)	gives the distribution function of Weibull distribution
3	qweibull3	qweibull3qweibull3(p, shape, scale=1, thres=0, lower.tail=TRUE, log.p=FALSE)	gives the quantile function of Weibull distribution
4	rweibull3	rweibull3(n, shape, scale=1, thres=0)	generates random deviates of Weibull distribution

3.2.2 weibullness

Performs a goodness-of-fit test of Weibull distribution (weibullness test) and provides the maximum likelihood estimates of the three-parameter Weibull distribution. Note that the threshold parameter is estimated based on the correlation from the Weibull plot. This package in R-software includes following commands:

Table 3.2 Commands used in weibullness package

Sr. No.	Syntax	Purpose
1	weibull.mle(x, threshold)	Maximum likelihood estimates of three-parameter Weibull distribution
2	weibull.threshold(x, a, interval.threshold, extendInt="downX")	Estimate of threshold parameter of three-parameter Weibull distribution by maximizing the correlation function from the Weibull plot
3	weibull.wp(x, n, a=0.5)	Estimate of shape and scale parameters of Weibull distribution using the intercept and slope estimates from the Weibull plot
4	wp.test(x, a)	Performs the Weibullness Test using the sample correlation from the Weibull Plot
5	wp.test.critical(alpha, n)	Calculates the critical value for the Weibullness test
6	wp.test.pvalue(r, n)	Calculates the p-value for the Weibullness test based on the sample correlation from the Weibull plot
7	Weibull.Plot.Quantiles	Plots Weibull quantile values

3.2.3 FITDISTRPLUS

The package fitdistrplus provides functions for fitting univariate distributions to different types of data (continuous censored or non-censored data and discrete data) and allowing different estimation methods (maximum likelihood, moment matching, quantile matching and maximum goodness-of-fit estimation).This package also provides various functions to compare the fit of several distributions to a same data set and can handle bootstrap of parameter estimates. This package includes following commands:

- FITDIST

  Syntax: fitdist(data, distr, method=c("mle", "mme", "qme", "mge"), start=NULL, fix.arg=NULL, discrete, keepdata=TRUE, keepdata.nb=100, ...)

Description: This command is used to fit univariate distributions to non-censored data by maximum likelihood (mle), moment matching (mme), quantile matching (qme) or maximizing goodness-of-fit estimation (mge).

- GOFSTAT

  Syntax: gofstat(f, chisqbreaks, meancount, discrete, fitnames=NULL)

  Description: It computes the goodness-of-fit statistics for parametric distributions fitted to a same non-censored data set.

- PLOTDIST

  Syntax: plotdist(data, distr, para, histo =TRUE, breaks = "default", demp = FALSE, discrete, ...)

  Description: In order to plot an empirical distribution (non-censored data) with a theoretical one if specified, then this command is used.

3.2.4 WeibullFit

Provides a single function to fit data of an input data frame into one of the selected Weibull functions (w2, w3 and it's truncated versions), calculating the scale, location and shape parameters accordingly. This package include following Syntax:

- weibullFit

  Syntax: weibullFit(dataFrame, primaryGroup="parcela", secondaryGroup="idadearred", restrValue, pValue="dap", leftTrunc=5, folder=NA, limit=1e+05, selectedFunctions=NULL, amp=2, pmaxIT=20, verbose=FALSE)

  Description: It calculates the weibull function scale, shape and location parameters using the maximum-likelihood method. The resulting plots and files are saved into the 'folder' parameter provided by the user.

3.2.5 ForestFit

The command 'fitWeibull' is ues to estimate the parameters of the two-parameter and three-parameter Weibull model. This package includes following command:

- fitWeibull
  Syntax: fitWeibull(data, location, method, starts)
  Description: In three-parameter case the methods for estimating parameters are :
- "mle" (for the method of ML)
- "moment" (for the method of moment
- "mps" (for the method of maximum product spacing)
- "wml" (for the method of weighted ML).
  A list of objects in two parts is given by the following command:
  1. Estimated parameters for three-parameter Weibull distribution.
  2. A sequence of goodness-of-fit measures consist of Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (AD), Crameer-von Misses (CVM), Kolmogorov-Smirnov (KS), and log-likelihood (log-likelihood) statistics.

Chapter 4

Estimation and Simulation Study for Weibull Distribution

4.1 Parameter Estimation

Parameter estimation is concerned with finding the value of a population parameter from sample statistics. In general, there are many methods to estimate the parameters of a distribution, such as probability-weighted moment, maximum likelihood method, least square estimates and so on. The maximum likelihood estimate is the most widely used method of parameter estimation. In this section we are going to study the parameter estimation methods such as maximum likelihood method and method of moments.

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample of size n drawn from WD(μ, α, β) and $\bar{x}$ be the sample mean. We discuss below the estimates of α and β using the method of maximum likelihood and methods of moments based on a random sample. Furthermore, we compare the performance of these estimators using simulation.

4.1.1 Maximum Likelihood Estimators

The likelihood function based on the random sample is the joint density of the n random variables and is a function of the unknown parameter. Thus, the likelihood function is:

$$L = \prod_{i=1}^{n} f_{x_i}(x_i, \alpha, \beta)$$

Now, we apply the MLE method to estimate the Weibull parameters, namely the shape parameter and the scale parameter. Consider the Weibull probability density function (pdf) given in the likelihood function will be

$$L(x_1, x_2, ...x_n; \alpha, \beta) = \prod_{i=1}^{n} \frac{\alpha}{\beta}\left(\frac{x_i - \mu}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x_i - \mu}{\beta}\right)^{\alpha}}$$

Here we assume that $\mu = 0$. Differentiating log-likelihood equation with respect to α and β in turn and equating it to zero, we obtain the estimating equations as:

$$\frac{\partial log L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} ln\, x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i^{\alpha} ln\, x_i = 0$$

$$\frac{\partial log L}{\partial \beta} = -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i^{\alpha} = 0$$

On elimination of β between these two equations and simplifying, we have,

$$\frac{\sum_{i=0}^{n} ln\, x_i}{\sum_{i=1}^{n} x_i^{\alpha}} - \frac{1}{\alpha} - \frac{1}{n}\sum_{i=1}^{n} ln\, x_i = 0$$

which may be solved to get the estimate of $\hat{\alpha}$. This can be accomplished by the use of standard iterative procedures (i.e., Newton-Raphson

method). Once $\hat{\alpha}$ is determined, $\hat{\beta}$ can be estimated using equation as given below:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i^{\hat{\alpha}}}{n}$$

4.1.2 Method of Moments

The method of moments is another technique commonly used in the field of parameter estimation. If the numbers $x_1, x_2, ...x_n$ represent a set of data, then an unbiased estimator for the k^{th} origin moment is

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

where; $\hat{m}_k$ stands for the estimate of m_k . In Weibull distribution, the k^{th} moment readily follows:

$$m_k = \frac{1}{\beta^{\alpha}}^{\frac{k}{\alpha}} \Gamma(1+\frac{k}{\alpha}) \tag{4.1}$$

where Γ signifies the gamma function as:

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx, \quad (s>0)$$

Then from equation (1), we can find the first and the second moment as follows

$$m_1 = \left(\frac{1}{\beta}\right)^{\frac{1}{\alpha}} \Gamma\left(1+\frac{1}{\alpha}\right)$$

$$m_2 = \left(\frac{1}{\beta}\right)^{\frac{2}{\alpha}} \left\{\Gamma\left(1+\frac{2}{\alpha}\right) - \left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}$$

When we divide m_2 by the square of m_1 , we get an expression which is a function of α only

$$\frac{m_2}{m_1^2} = \frac{\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})}{\Gamma^2(1+\frac{1}{\alpha})}$$

on taking the square roots of above equation, we have coefficient of variation

$$CV = \frac{\sqrt{\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})}}{\Gamma(1+\frac{1}{\alpha})}$$

Now, we can form a table for various CV by using above equation for different α values. In order to estimate α and β , we need to calculate the coefficient of variation CV_d of the sample data. Having done this, we compare CV_d with CV using the table. The corresponding α is the estimated one $\widehat{\alpha}$. The β can be estimated using the following

$$\hat{\beta} = \left[\frac{\bar{x}}{\Gamma(1+\frac{1}{\hat{\alpha}})}\right]^{\hat{\alpha}}$$

where $\bar{x}$ is the mean of the data.

4.2 A Simulation study for studying the performance of the estimators

The comparability of the two methods of estimation (Maximum Likelihood method, Method of Moments) is explored via simulation study involving fixed value of threshold parameter($\mu = 3$), shape parameter ($\alpha = 1.3$) and scale parameter ($\beta = 2$) and various sample sizes ranging from 5 to 50. Here we study the performance of MLE and MOM with respect to sample size n, and both the parameters μ and β are fixed. Following are the steps considered for the study:

1. Generate 1000 samples of size n.
2. For each sample we get 1000 estimates of parameters of MLE and MOM for μ, α and β.
3. Calculate MSE for each MLE and MOM of μ, α and β based on 1000 estimates of parameters. These are denoted as $MSE(\hat{\mu}_{MLE})$ and $MSE(\hat{\mu}_{MOM})$ respectively.

4. Calculate Bias for each MLE and MOM of μ, α and β based on 1000 estimates of parameters. Take average of the bias calculated. These are denoted as $Bias(\hat{\mu}_{MLE})$ and $Bias(\hat{\mu}_{MOM})$.
5. The above steps are repeated for different values of $n = 5, 10, \ldots, 50$.

The R-program to implement the above algorithm is given below.

R-code to calculate MSE and Bias for the simulated data

```
library(ForestFit)
library(FAdist)
alpha=1.3;beta=2.4;mu=3; n=5;
d=rweibull3(n*1000,shape=alpha,scale=beta,thres=mu)
x=matrix(d,nrow=n,ncol=1000)
x=data.frame(x)
#————————————————————————————————
# ESTIMATING PARAMETER USING MLE
#————————————————————————————————
X=function(x){
starts=c(1.3,2.4,3)
z=fitWeibull(x,1,"ml",starts)$estimate[1,]
}
y1=apply(x,2,X)
typeof(y)
y1=unlist(y1, use.names = 0)
alpha_hat=y1[1,]
alpha_hat_mle=mean(alpha_hat)
alpha_hat_mle
mse_alpha_mle=mean((round(alpha_hat,digits=5)-alpha)^2)
mse_alpha_mle
bias_alpha_mle=mean(round(alpha_hat, digits = 5))-alpha
bias_alpha_mle
```

```
beta_hat=y1[2,]
beta_hat_mle=mean(beta_hat)
beta_hat_mle
mse_beta_mle=mean((round(beta_hat,digits = 5)-beta)^2)
mse_beta_mle
bias_beta_mle=mean(round(beta_hat,digits = 5))-beta
bias_beta_mle

mu_hat=y1[3,]
mu_hat_mle=mean(mu_hat)
mu_hat_mle
mse_mu_mle=mean((round(mu_hat,digits = 5)-mu)^2)
mse_mu_mle
bias_mu_mle=mean(round(mu_hat,digits = 5))-mu
bias_mu_mle
#-------------------------------------------------
# ESTIMATING PARAMETER USING MOM
#-------------------------------------------------
Y=function(x){
starts=c(1.3,2.4,3)
z=fitWeibull(x,1,"moment",starts)$estimate[1,]
}
y2=apply(x,2,Y)
typeof(y2)
y2=unlist(y2, use.names = 0)
alpha_hat_mom=mean(y2[1,])
alpha_hat_mom
mse_alpha_mom=mean((y2[1,]-alpha)^2)
mse_alpha_mom
bias_alpha_mom=mean(y2[1,])-alpha
bias_alpha_mom
```

```
beta_hat_mom=mean(y2[2,])
beta_hat_mom
mse_beta_mom=mean((y2[2,]-beta)^2)
mse_beta_mom
bias_beta_mom=mean(y2[2,])-beta
bias_beta_mom

mu_hat_mom=mean(y2[3,])
mu_hat_mom
mse_mu_mom=mean((y2[3,]-mu)^2)
mse_mu_mom
bias_mu_mom=mean(y2[3,])-mu
bias_mu_mom
#-------------------------------------------------

n=c(5,10,15,20,25,30,35,40,45,50)

cbind(n,alpha_hat_mle,alpha_hat_mom, mse_alpha_mle,
mse_alpha_mom,bias_alpha_mle, bias_alpha_mom)

cbind(n, beta_hat_mle, beta_hat_mom, mse_beta_mle,mse_beta_mom,
bias_beta_mle, bias_beta_mom)

cbind(n, mu_hat_mle, mu_hat_mom, mse_mu_mle,
mse_mu_mom,bias_mu_mle, bias_mu_mom)
#-------------------------------------------------
# Plotting of MSE for threshold parameter
#-------------------------------------------------
plot(n,mse_mu_mle,pch=18,cex=2,lwd=2.5,"o",xlab="n",ylab="MSE",
main = "MSE of threshold estimators",col="dark blue")
```

```
points(n,mse_mu_mom,pch=20,cex=2,lwd=2.5,"o",col="red")
legend(x="topright",legend=c("mu_mle","mu_mom"), col=c("dark blue","red"),lty=1)

#————————————————————————————————
# Plotting of MSE for shape parameter
#————————————————————————————————
plot(n,mse_alpha_mle,pch=18,cex=2,lwd=2.5,"o",xlab="n",ylab="MSE", main = "MSE of shape estimators",col="dark blue")

points(n,mse_alpha_mom,pch=20,cex=2,lwd=2.5,"o",col="red")
legend(x="topright",legend=c("alpha_mle","alpha_mom"), col=c("dark blue","red"),lty=1)
#————————————————————————————————
# Plotting of MSE for scale parameter
#————————————————————————————————
plot(n,mse_beta_mle,pch=18,cex=2,lwd=2.5,xlab="n",ylab="MSE", main = "MSE of scale estimators",col="dark blue")

points(n,mse_beta_mom,pch=20,cex=2,lwd=2.5,"o",col="red")
legend(x="bottomleft",legend=c("beta_mle","beta_mom"), col=c("dark blue","red"),lty=1)
```

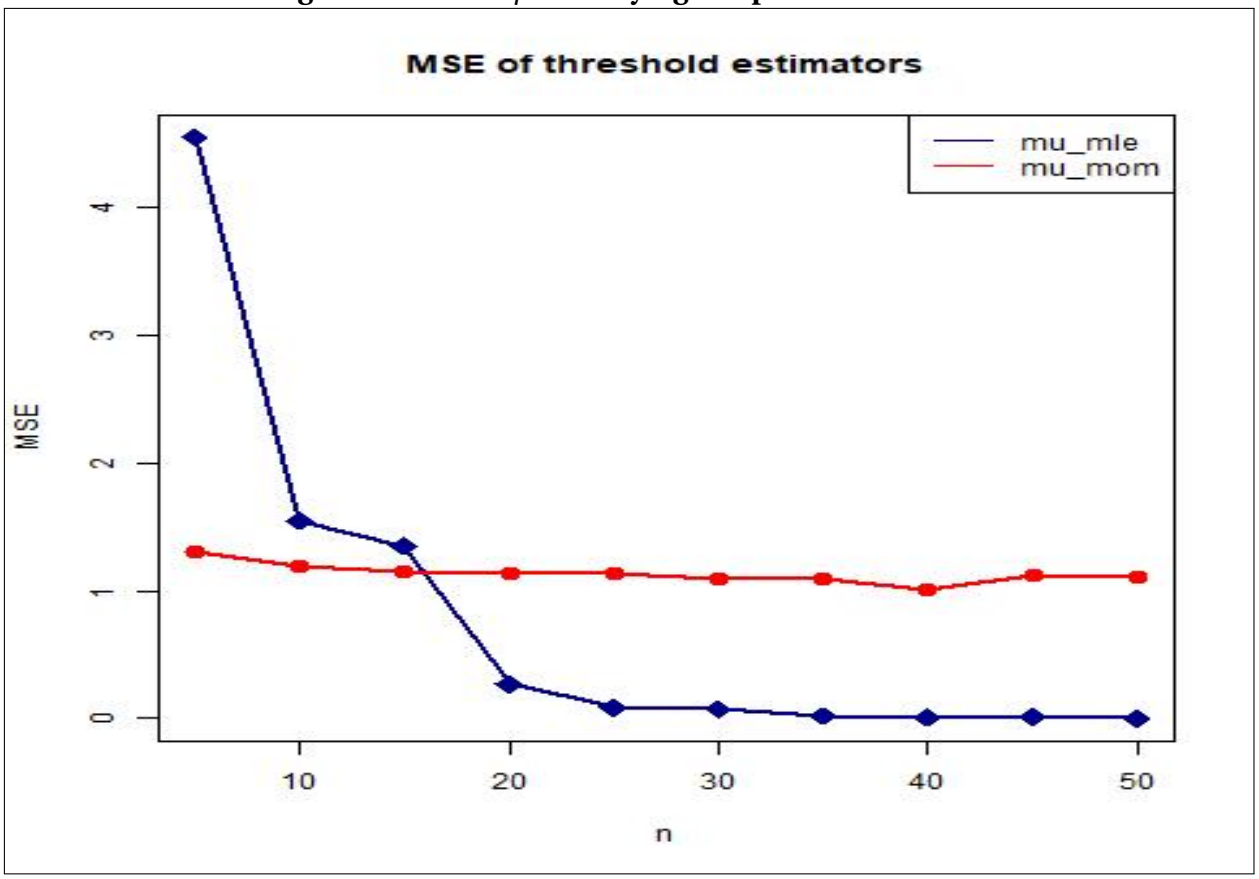
The results obtained for $\hat{\mu}$are as follows in Table 4.1.

Table 4.1 Table of MSE and Bias for $\hat{\mu}$

n	*MLE*	*MOM*	**MSE**		**Bias**	
			MLE	*MOM*	*MLE*	*MOM*
5	1.50573	4.03734	4.54464	1.29773	-1.49414	1.03734
10	2.33742	4.03986	1.55097	1.19215	-0.66258	1.03986
15	2.48966	4.03416	1.34900	1.14328	-0.51034	1.03416
20	2.78336	4.04228	0.26743	1.14079	-0.21664	1.04228
25	2.84831	4.04236	0.08625	1.12859	-0.15166	1.04536
30	2.89061	4.03126	0.07299	1.09654	-0.1094	1.03126
35	2.92646	4.03269	0.01956	1.0936	-0.07356	1.03269
40	2.94875	4.03786	0.0084	1.00131	-0.05123	1.03786
45	2.95207	4.04624	0.00842	1.11676	-0.04792	1.04624
50	2.96425	4.04044	0.00327	1.10166	-0.03576	1.04044

We plot the MSE of $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\beta}$ estimators across sample size n.

Figure 4.1 MSE of $\hat{\mu}$ for varying sample size n.



Observations: In order to study the performance of μ-estimators in terms of MSE for varying n and fixed μ, α and β we observe from Figure 4.1,

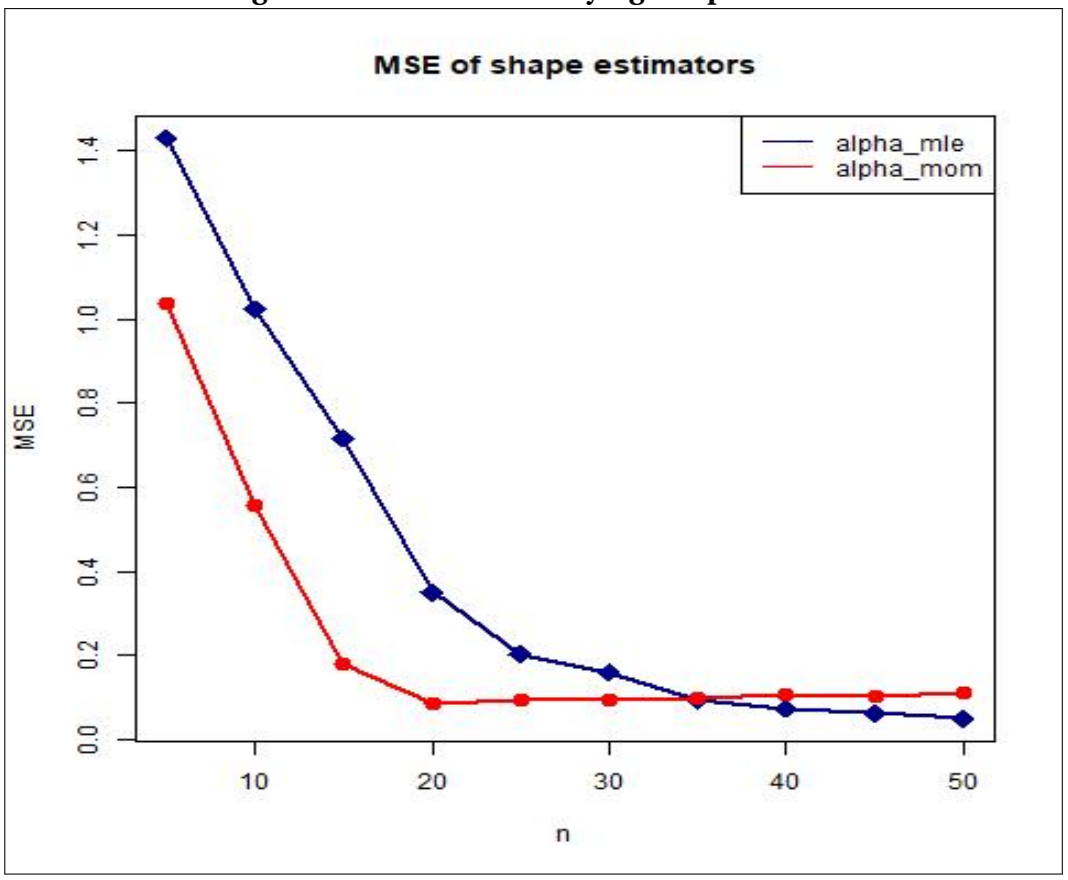
1. In general as n increases MSE of all estimators decreases.
2. Here, MSE($\hat{\mu}_{MLE}$) > MSE($\hat{\mu}_{MOM}$) i.e. MSE($\hat{\mu}_{MLE}$) is maximum than that of MSE($\hat{\mu}_{MOM}$).
3. For higher sample sizes i.e. $n \geq 20$, MSE($\hat{\mu}_{MLE}$) stabilizes gradually.
4. There is no effect on MSE($\hat{\mu}_{MOM}$) as sample size increases.

The results obtained for $\hat{\alpha}$ are as follows in Table 4.2.

Table 4.2 Table of MSE and Bias for $\hat{\alpha}$

n	*MLE*	*MOM*	**MSE**		**Bias**	
			MLE	*MOM*	*MLE*	*MOM*
5	0.78508	1.12102	1.43005	1.03712	-0.51493	-0.17898
10	1.04891	1.06557	1.02262	0.55592	-0.25109	-0.23443
15	1.11489	1.03485	0.71744	0.17780	-0.18511	-0.26515
20	1.13812	1.0194	0.34907	0.08574	-0.16189	-0.2806
25	1.15543	1.00836	0.20161	0.09164	-0.14458	-0.29164
30	1.17544	1.0022	0.15901	0.09448	-0.12457	-0.2978
35	1.18932	0.99587	0.0921	0.09830	-0.11068	-0.30413
40	1.20545	0.98512	0.07191	0.10482	-0.09455	-0.31488
45	1.22305	0.98695	0.06175	0.10278	-0.07695	-0.31305
50	1.22454	0.97784	0.04979	0.10887	-0.07547	-0.32216

Figure 4.2 MSE of $\hat{\alpha}$ for varying sample size n.



Observations: In order to study the performance of α-estimators in terms of MSE for varying n and fixed μ, α and β we observe from Figure 4.2,

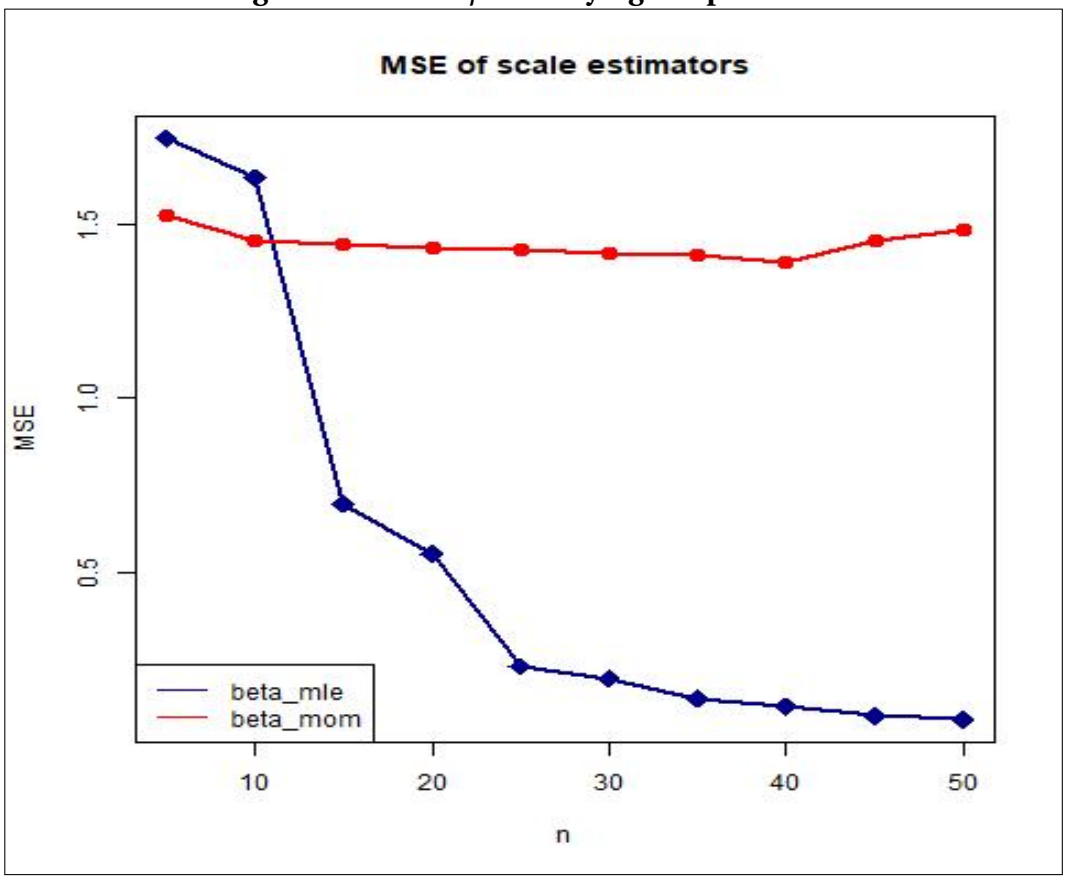
1. In general as n increases MSE of all estimators decreases.
2. Here, MSE($\hat{\alpha}_{MOM}$) < MSE($\hat{\alpha}_{MLE}$) i.e. MSE($\hat{\alpha}_{MLE}$) is maximum than that of MSE($\hat{\alpha}_{MOM}$).
3. For higher sample sizes i.e. $n \geq 30$, MSE($\hat{\alpha}_{MLE}$) stabilizes gradually.

The results obtained for $\hat{\beta}$are as follows in Table 4.3.

Table 4.3 Table of MSE and Bias for $\hat{\beta}$

n	*MLE*	*MOM*	**MSE**		**Bias**	
			MLE	*MOM*	*MLE*	*MOM*
5	1.92101	1.81617	1.74415	1.52412	-0.07899	-1.18383
10	1.84689	0.81992	1.63141	1.45359	-0.15311	-1.18008
15	1.82148	0.81581	0.69713	1.43964	-0.17852	-1.18419
20	1.84603	0.81567	0.55124	1.43178	-0.15398	-1.18433
25	1.83251	0.81592	0.22932	1.42594	-0.16748	-1.18408
30	1.82003	0.80068	0.19525	1.41707	-0.17998	-1.19932
35	1.85459	0.80315	0.13722	1.40859	-0.1454	-1.19685
40	1.86345	0.79402	0.11334	1.388	-0.13656	-1.20598
45	1.89299	0.80057	0.08782	1.451	-0.10701	-1.19943
50	1.87356	0.78752	0.07842	1.48061	-0.12644	-1.21248

Figure 4.3 MSE of $\hat{\beta}$ for varying sample size n.



Observations: In order to study the performance of β-estimators in terms of MSE for varying n and fixed μ, α and β we observe from Figure

4.3,

1. For MSE($\hat{\beta}_{MLE}$) as n increases MSE of all estimators decreases.
2. MSE($\hat{\beta}_{MOM}$) isn't diminishing much.
3. MSE($\hat{\beta}_{MLE}$) < MSE($\hat{\beta}_{MOM}$)
4. For higher sample sizes i.e. $n \geq 30$, MSE($\hat{\beta}_{MLE}$) stabilizes gradually.

4.3 Conclusions

In this chapter, two methods for estimating the parameters of the Weibull distribution: Maximum Likelihood estimators (MLE) and Method of Moments (MOM) are described. The performance of these methods is compared using a simulation study. The efficiency of the methods is compared on the basis of the MSE criterion and sample size n. As the sample size n increases, the MSE of both methods decreases and hence the estimation precision of the parameter increases.

It is evident that MLE performs better than MOM when the sample size is medium or larger than enough. Both methods are good enough for their simplicity. There is one complication while using MOM. This method needs to use the gamma function. However, the gamma function can be easily obtained by using the software Matlab. The performance of MLE is often better than MOM. The MLE is the most popular because of its efficiency, good properties and it is simpler to compute with respect to MOM. Therefore, we recommend the MLE method to estimate the parameters of the Weibull distribution.

Chapter 5

Fitting of Weibull Distribution

5.1 Fitting of Weibull Distribution to Birnbaum and Saunders (1969) data

Description of data: The data contains the lifetime data in this case correspond to the cycles ($\times 10^{-3}$) of aluminum specimens of type $6061 - T6$. They were exposed to a pressure with maximum stress of $21,000 psi$. All specimens were tested until failure. Table 5.1 presents the data.

Structure of data:

- Data is given by: Birnbaum and Saunders (1969)
- Variable under study: Lifetime (x_i) of aluminium specimens
- $n = 101$

Descriptive statistics of data

```
       xi
Min.   : 370
1st Qu.:1115
Median :1416
Mean   :1401
3rd Qu.:1642
Max.   :2440
```

Table 5.1 Birnbaum and Saunders data (1969)

Sr. No.	x_i	**Sr. No.**	x_i	**Sr. No.**	x_i	**Sr. No.**	x_i	**Sr. No.**	x_i
1	370	21	1055	41	1270	61	1502	81	1763
2	706	22	1085	42	1290	62	1505	82	1768
3	716	23	1102	43	1293	63	1513	83	1781
4	746	24	1102	44	1300	64	1522	84	1782
5	785	25	1108	45	1310	65	1522	85	1792
6	797	26	1115	46	1313	66	1530	86	1820
7	844	27	1120	47	1315	67	1540	87	1868
8	855	28	1134	48	1330	68	1560	88	1881
9	858	29	1140	49	1355	69	1567	89	1890
10	886	30	1199	50	1390	70	1578	90	1893
11	886	31	1200	51	1416	71	1594	91	1895
12	930	32	1200	52	1419	72	1602	92	1910
13	960	33	1203	53	1420	73	1604	93	1923
14	988	34	1222	54	1420	74	1608	94	1924
15	990	35	1235	55	1450	75	1630	95	1945
16	1000	36	1238	56	1452	76	1642	96	2023
17	1010	37	1252	57	1475	77	1674	97	2100
18	1016	38	1258	58	1478	78	1730	98	2130
19	1018	39	1262	59	1481	79	1750	99	2215
20	1020	40	1269	60	1485	80	1750	100	2268
								101	2440

For the estimating the parameters of We fit Weibull distribution to the data using ForestFit package in R-software.

R-code to fit Weibull distribution to Birnbaum and Saunders data

```
library(ForestFit)
data=read.csv("saunders data.csv")
data=data[2];data
starts=c(1,1357,180)
fitWeibull(data$xi,1 ,"ml", starts)
```

The following is the output of fitting Weibull distribution to Birnbaum and Saunders data obtained from R-software using ForestFit package.

R-output of fitting Weibull distribution to Birnbaum & Saunders data

```
$estimate
        alpha     beta       mu
[1,] 3.429694 1356.191 181.1529

$measures
        AIC     CAIC      BIC     HQIC        AD
[1,] 1497.4   1497.647 1505.245  1500.576  0.1942208
        CVM          KS       log.likelihood
[1,] 0.02606791  0.04744844     -745.6999
```

Now we fit Weibull distribution to the data using Weibullness package in R.

R-code to fit Weibull distribution to Birnbaum and Saunders data

```
library(weibullness)
data=read.csv("saunders data.csv")
data=data[2];data
y=weibull.threshold(data$xi, 0.91)
print(y) print(weibull.mle(data$xi, y))
print(wp.test(data$xi, 0.91))
```

The following is the output of fitting Weibull distribution to Birnbaum and Saunders data obtained from R-software using weibullness package.

R-output of fitting Weibull distribution to Birnbaum & Saunders data

```
threshold
182.6804

shape     scale  threshold
3.425052  1354.615   182.6804

Weibullness test from the Weibull plot

data:  data$xi
correlation = 0.99099, p-value = 0.3775
```

We fit Weibull distribution to Birnbaum and Saunders data using fitdistrplus package in R. This package fit Weibull distribution with 2-parameters. The R-code for fitting the distribution is as follows.

R-code to fit Weibull distribution to Birnbaum and Saunders data

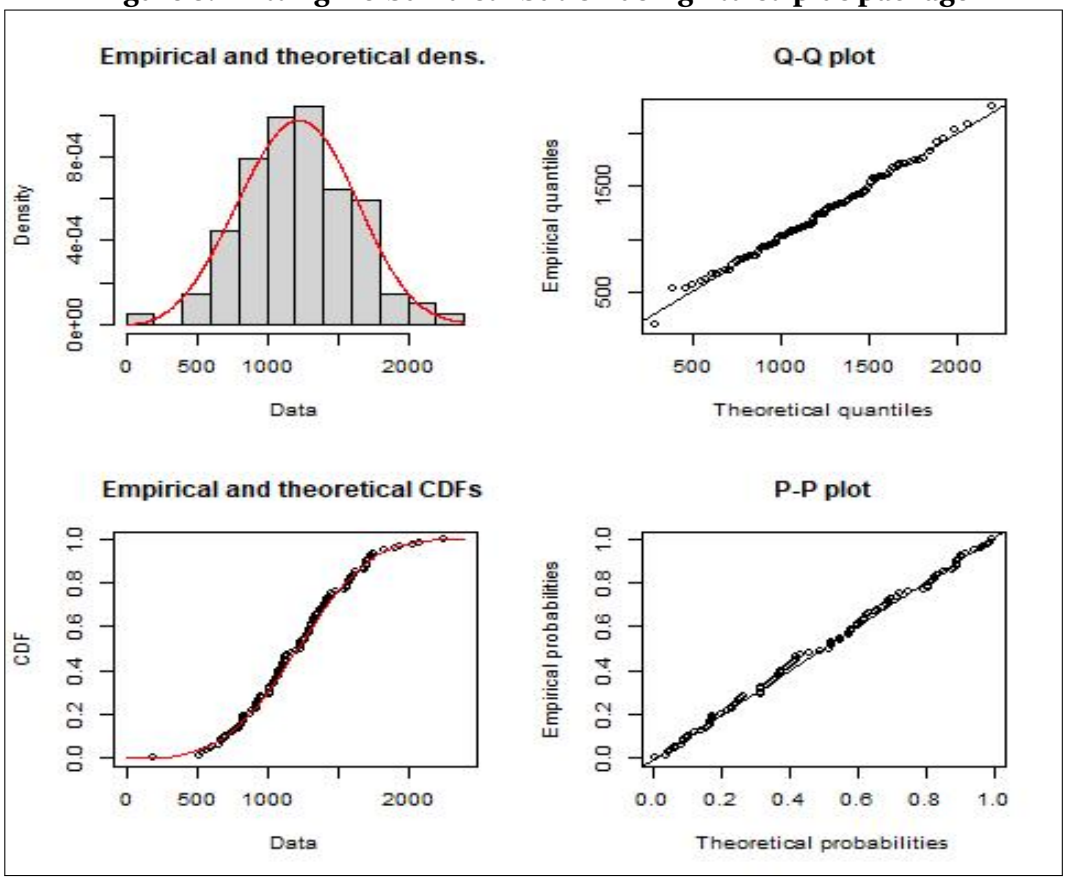
```
library(fitdistrplus)
fw=fitdist(data$Value-y, "weibull")
summary(fw)
plot(fw)
```

The following is the output of fitting Weibull distribution to Birnbaum and Saunders data obtained from R-software using fitdistrplus package.

R-output of fitting Weibull distribution to Birnbaum & Saunders data

```
Fitting of the distribution ' weibull ' by maximum likelihood
Parameters :
       estimate     Std. Error
shape 3.429716 0.2636959
scale 1356.332282 41.4760837
Loglikelihood:  -745.6999   AIC:  1495.4   BIC:  1500.63
Correlation matrix:
          shape     scale
shape 1.0000000 0.3163862
scale 0.3163862 1.0000000
```

Figure 5.1 Fitting Weibull distribution using fitdistrplus package



5.2 Fitting of Weibull Distribution to Annual Peak Stream Flow (cfs) (1981 to 2015) Data

The data contains the value recorded by United States Geological Survey's (USGS) annual peak streamflow recorded at a USGS gauging station in the United States. Generally a minimum time period of 30 years is considered ideal for flood frequency analysis and therefore, the annual peak streamflow data from 1981 to 2015 is used.

Structure of data:

- Experiment counducted at: Wabash River at Lafayette, Indiana
- Variable under study: Annual peak stream flow (y_i) in cubic feet per second(cfs)
- $n = 35$

Table 5.2 Annual Peak Stream Flow (cfs) (1981 to 2015) Data

Year	y_i	**Year**	y_i	**Year**	y_i	**Year**	y_i	**Year**	y_i
1981	44500	1988	33300	1995	30800	2002	40700	2009	59200
1982	56400	1989	40700	1996	35400	2003	80000	2010	44300
1983	60800	1990	53300	1997	54800	2004	58300	2011	58400
1984	40400	1991	77400	1998	54500	2005	80000	2012	45800
1985	80400	1992	33300	1999	61100	2006	30100	2013	85700
1986	41600	1993	62500	2000	31000	2007	50800	2014	51500
1987	14700	1994	65600	2001	34800	2008	72400	2015	69500

Descriptive statistics of data

```
     stream
Min.   :14700
1st Qu.:40550
Median :53300
Mean   :52400
3rd Qu.:61800
Max.   :85700
```

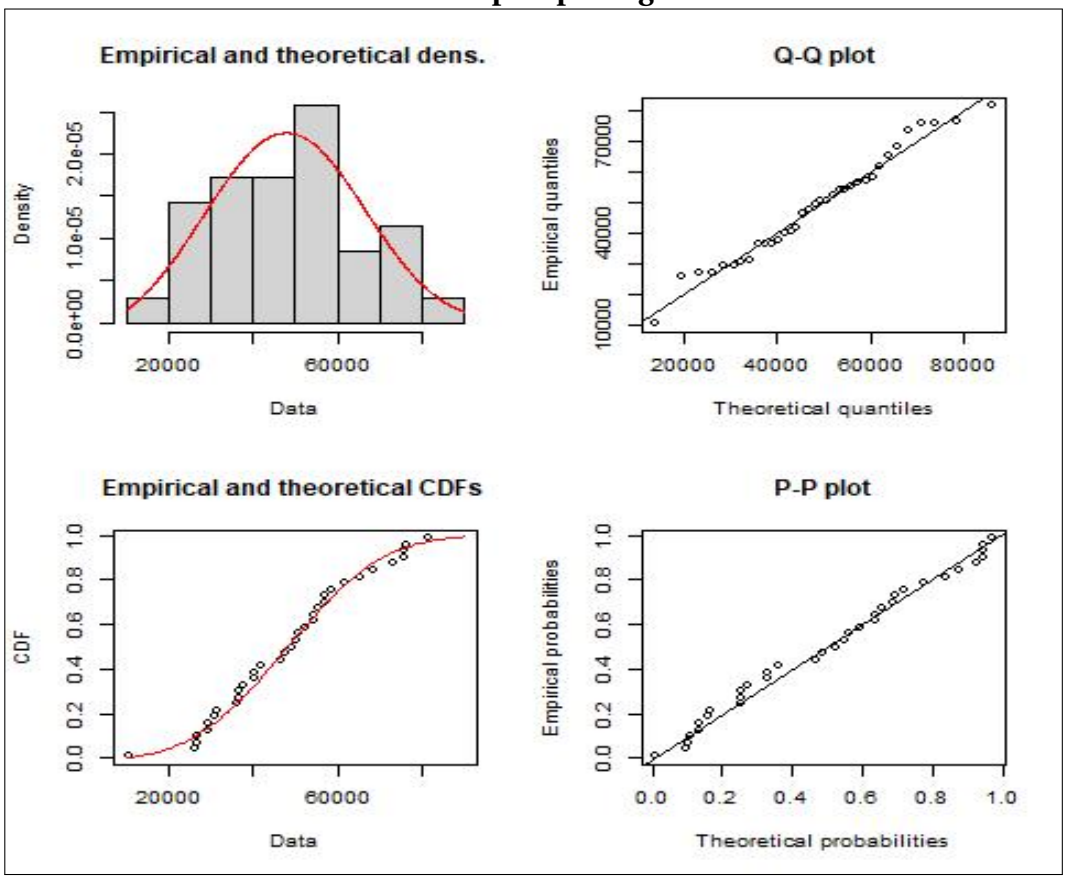
The fits to the Annual Peak Stream Flow (cfs) data are given in the following table using ForestFit and weibullness packages in R-software.

Table 5.3 The fits to the Annual Peak Stream Flow (cfs) Data w.r.t packages in R-softwa

Sr. No.	**Package**	**Method**	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	**AIC**	**BIC**	**p-value**
1	ForestFit	MLE	3973.935	3.1257	54162.96	786.8623	791.5284	NA
2	ForestFit	MPS	-6606.995	3.5312	65557.42	787.65	792.3161	NA
3	Weibullness	MLE	3958.4	3.1270	54184.96	NA	NA	0.5165

The following is the output of fitting Weibull distribution to Annual Peak Stream Flow (cfs) data using fitdistrplus package in R.

Figure 5.2 Fitting Weibull distribution to Annual Peak Stream Flow (cfs) using fitdistrplus package



5.3 Fitting of Weibull Distribution for simulated data

We have Fit the Weibull Distribution for real life data in section 5.1 and 5.2. Now let's see how we can fit the Weibull Distribution on simulated data. We can simulate the data from weibull distribution using FAdist package and fit this simulated data using ForestFit package in R. The Table 5.3 shows the simulated data.

Structure of data:

- Experiment: Simulated data
- Variable under study: Values of random sample (x_i)
- $n = 1000$
- Distribution of data: $W(\mu = 5, \alpha = 2, \beta = 1.5)$

Table 5.3 Simulated data from $Weibull(\mu = 5, \alpha = 2, \beta = 1.5)$

6.0352	6.4137	7.0932	5.7105	5.5437	5.9596
5.5003	5.9388	6.2764	6.9145	7.0894	7.2736
7.1419	6.8036	6.9393	7.1398	6.2194	6.0165
5.5211	6.2155	5.5993	6.3045	5.2843	6.2067
6.5512	7.028	5.7722	6.5768	5.1744	6.4669
6.1531	5.6968	6.8572	5.3946	5.601	7.1066
6.9514	6.4319	6.0988	8.6505	6.1661	5.9111
5.2157	5.9604	5.7757	7.7894	5.1284	6.0012
5.7167	5.6832	5.58	5.9853	5.9382	5.3912
6.8676	5.5244	7.2466	6.1806	6.167	6.1361
6.5438	6.9935	6.4499	5.9735	6.6098	5.5427
5.4293	6.5687	6.4114	8.2573	5.4136	5.8118
5.4414	6.4367	5.6513	7.2776	6.4986	6.9385
5.5577	6.387	6.3224	6.7671	6.8406	6.2689
5.861	6.3823	6.3593	6.8633	5.6371	5.7302
6.7081	6.9287	5.8182	6.4524	6.6563	7.3596
5.4828	7.2753	5.3128	7.1162	5.0367	6.2165
7.0436	6.3863	7.0343	7.497	5.8293	7.1227

Descriptive statistics of data

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.037   5.786   6.237   6.327   6.806   9.252
```

The following data is generated from FAdist package using the following command.

R-code to simulate data from Weibull distribution

```
library(FAdist)
n=1000; alpha=2;
mu=5; beta=1.5;
x=rweibull3(n,alpha,beta,mu)
x=round(x, digits=4)
x
```

The following is R-code to obtain estimates of parameters from various methods using ForestFit package.

R-code to simulate and fit the data from Weibull distribution

```
library(ForestFit)
starts=c(2,1,1.5)
fitWeibull(x,1 ,"ml", starts)
fitWeibull(x,1 ,"mps", starts)
```

The R-output to the simulated data using ForestFit package is given below.

R-output of fitting Weibull distribution to simulated data

```
$estimate
        alpha     beta       mu
[1,] 1.972268 1.491283 5.005542
$measures
          AIC     CAIC      BIC     HQIC        AD
[1,] 2020.095 2020.119 2034.818 2025.691 0.4224922
          CVM     KS        log.likelihood
[1,] 0.05926332 0.02050647      -1007.047
```

```
$estimate
        alpha     beta    mu
[1,] 1.987558 1.509048 4.991
$measures
          AIC     CAIC      BIC     HQIC        AD
[1,] 2020.549 2020.573 2035.272 2026.144 0.3798277
          CVM     KS        log.likelihood
[1,] 0.05035141  0.01895414      -1007.274
```

The following is the code for fitting Weibull distribution to simulated data using weibullness package in R-software.

R-code to simulate and fit the data from Weibull distribution

```
s = weibull.threshold(x, 0.809)
s
weibull.mle(x,s)
```

The following is the output of fitting Weibull distribution to simulated data using weibullness package in R.

R-output of fitting Weibull distribution to simulated data

```
threshold
5.010076


shape     scale threshold
 1.963273  1.485885  5.010076
```

The following is the code for fitting Weibull distribution to simulated data using fitdistrplus package in R.

R-code to simulate and fit the data from Weibull distribution

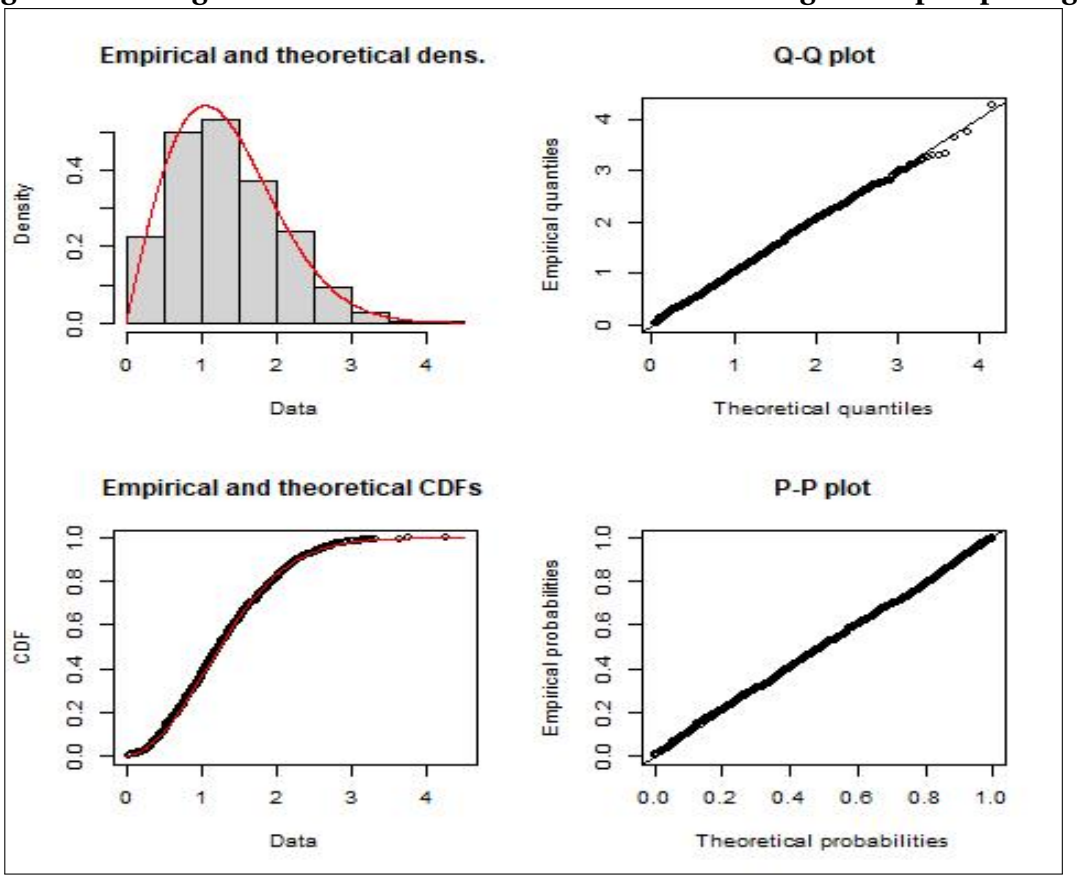
```
x
mu=5
fw=fitdist(x-mu, "weibull")
summary(fw)
plot(fw)
```

The following is the output of fitting Weibull distribution to simulated data using fitdistrplus package in R.

R-output of fitting Weibull distribution to simulated data

```
Fitting of the distribution ' weibull ' by maximum likelihood
Parameters :
       estimate Std. Error
shape  1.983691 0.04926889
scale  1.498191 0.02514764
Loglikelihood:  -1007.076   AIC:  2018.153   BIC:  2027.968
Correlation matrix:
          shape     scale
shape 1.0000000 0.3128016
scale 0.3128016 1.0000000
```

Figure 5.3 Fitting Weibull distribution to simulated data using fitdistrplus package



Chapter 6

Survival Analysis of Weibull Distribution

6.1 Introduction

Survival analysis is a branch of statistics for analyzing the expected duration of time until one event occurs, such as death in biological organisms and failure in mechanical systems. This topic is also called reliability theory or reliability analysis in engineering, duration analysis or duration modelling in economics, and event history analysis in sociology. Survival analysis deals with predicting the time when a specific event is going to occur. It is also known as failure time analysis or analysis of time to death. Survival analysis is used to analyze data in which the time until the event is of interest. The response is often referred to as a **failure time**, **survival time**, or **event time**. For example,

- time until tumour recurrence
- Time until cardiovascular death after some treatment intervention
- Time until AIDS for HIV patients
- Time until a machine part fails
- The time from diagnosis of a disease until death.
- The time between administration of a vaccine and development of an infection.

- The time from the start of treatment of a symptomatic disease and the suppression of symptoms.

be used. – Time to event is restricted to be positive and has a skewed distribution. – The probability of surviving past a certain point in time may be of more interest than the expected time of event. – The hazard function, used for regression in survival analysis, can lend more insight into the failure mechanism than linear regression.

6.2 What is censoring ?

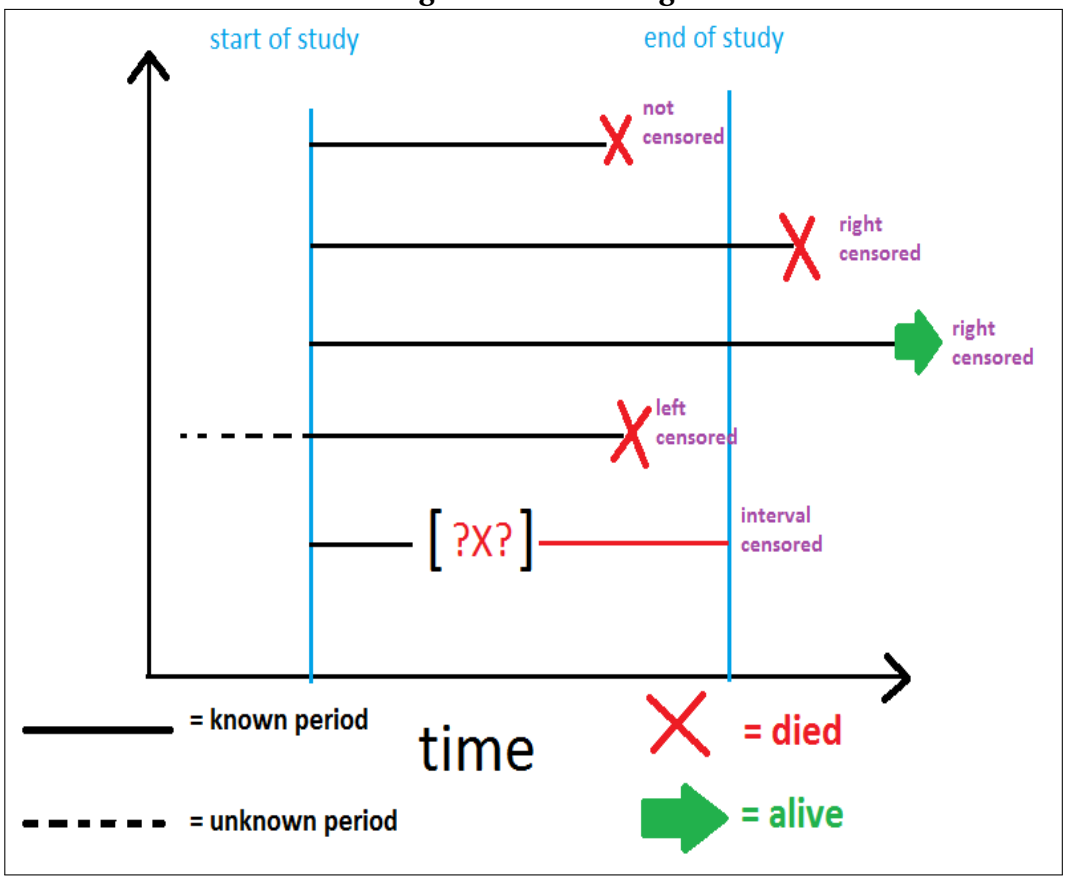
Censoring is a form of missing data problem in which time to event is not observed for reasons such as termination of study before all recruited subjects have shown the event of interest or the subject has left the study prior to experiencing an event. Censoring is common in survival analysis.

Censoring is present when we have some information about a subject's event time, but we don't know the exact event time. For the analysis methods we will discuss to be valid, censoring mechanism must be independent of the survival mechanism. There are generally three reasons why censoring might occur :

- A subject does not experience the event before the study ends
- A person is lost to follow-up during the study period
- A person withdraws from the study

These are all examples of right-censoring.

Figure 6.1 Censoring



6.3 Types of right censoring

- **Fixed type I censoring** occurs when a study is designed to end after C years of follow-up. In this case, everyone who does not have an event observed during the course of the study is censored at C years.
- In **random type I censoring** , the study is designed to end after C years, but censored subjects do not all have the same censoring time. This is the main type of right-censoring we will be concerned with.
- In **random type II censoring**, a study ends when there is a prespecified number of events.Regardless of the type of censoring, we must assume that it is non-informative about the event; that is, the censoring is caused by something other than the impending failure.

6.4 Survival function and Hazard function

Notation: T denotes the response variable where T>0.

The survival function is given as

$$S(t) = P(T > t) = 1 - F(t).$$

- The survival function gives the probability that a subject survives longer than time t.
- As t ranges from 0 to ∞, the survival function has the following properties
  - It is non-increasing
  - At time $t = 0$, $S(0) = 1$. In other words, the probability of surviving past time 0 is 1.
  - At time $t = \infty$, $S(\infty) = 0$. As time goes on increasing, the survival curve goes to 0.
- In theory, the survival function is smooth. In practice, we observe events on a discrete time scale(i.e. days, weeks, etc.).
- The hazard function, $h(t)$, is the instantaneous rate at which events occur, given no previous events.

$$h(t) = \lim_{\Delta \to 0} \frac{P(t < T\, t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$

- The cumulative hazard function describes the accumulated risk up to time t,

$$H(t) = \int_0^t h(u)\, du$$

- If we know any one of the functions $S(t)$, $H(t)$, or $h(t)$, we can derive the other two functions.

$$h(t) = \frac{\partial log(S(t))}{\partial t}, \quad H(t) = log(S(t)), \quad S(t) = exp(-H(t))$$

6.5 Survival Analysis for Weibull Distribution

Suppose we assume the time-to-event follows a WD(μ, α, β), then the survival function, hazard function and cumulative hazard function can be given as:

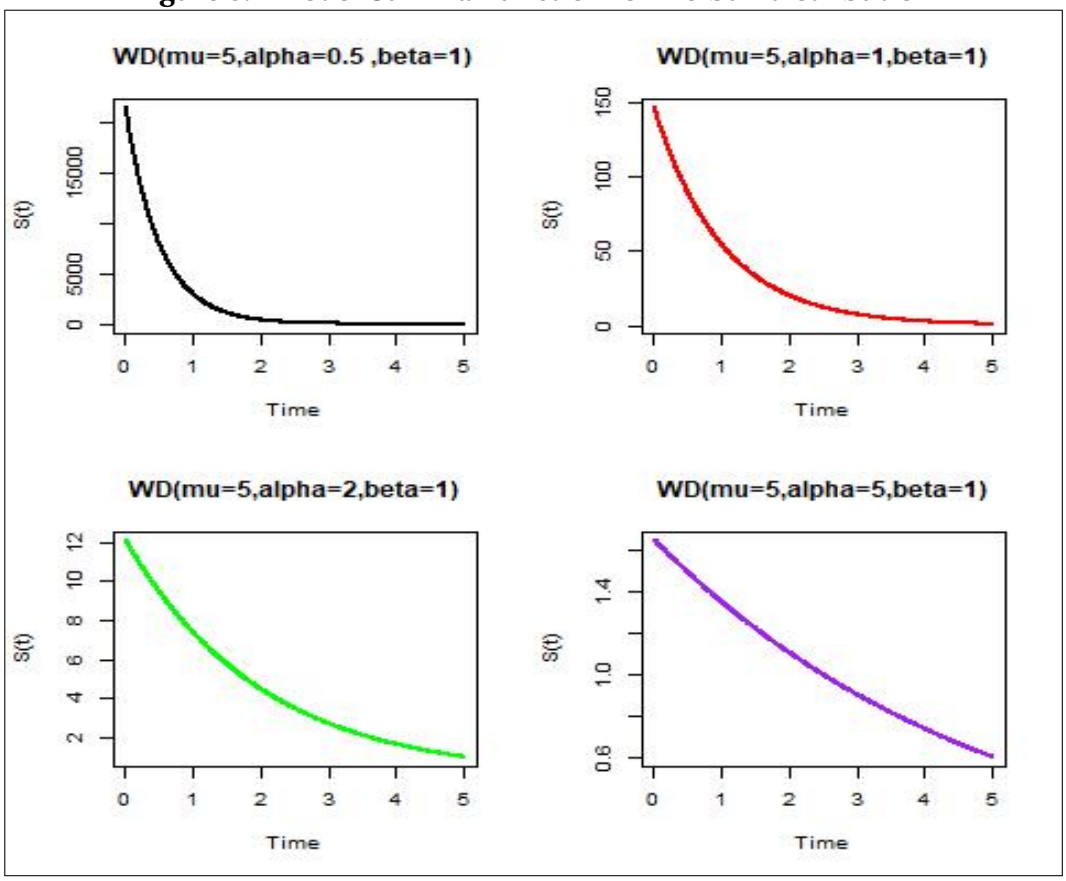
$$S(t) = exp\left[-\left(\frac{t-\mu}{\beta}\right)^{\alpha}\right]$$

$$h(t) = \frac{\alpha}{\beta}\left(\frac{t-\mu}{\beta}\right)^{\alpha}$$

$$H(t) = -\left(\frac{t-\mu}{\beta}\right)^{\alpha}$$

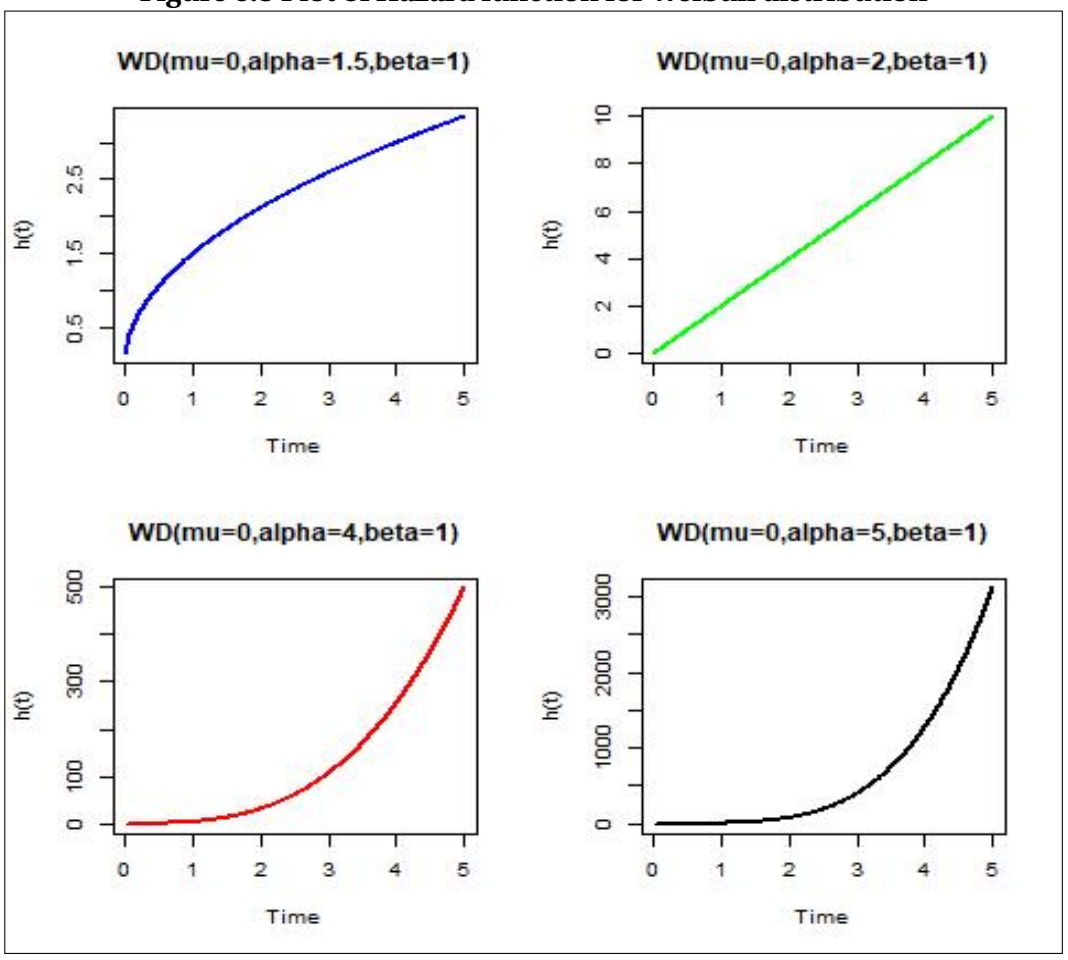
Following are the graphs of the Survival function and Hazard function for WD.

Figure 6.2 Plot of Survival function for Weibull distribution



Observations: As t increases from 0 to ∞, $S(t)$ increases if $\alpha > 0$ and decreases if $0 < \alpha < 1$. As α increases, $S(t)$ also increases.

Figure 6.3 Plot of Hazard function for Weibull distribution



Observations: As t increases from 0 to ∞, $h(t)$ increases if $\alpha > 1$ and decreases if $0 \leq \alpha < 1$. When $\alpha = 2$, we have the Rayeleigh distribution, and the hazard rate is the straight line passing through the origin. As time increases, failure rate also increases.

Note: No failure can occur before μ hours, so the time scale starts at μ, and not 0. If a shift parameter μ is known (based, perhaps, on the physics of the failure mode), then all you have to do is subtract μ from all the observed failure times or readout times and analyze the resulting shifted data with a two-parameter Weibull.

6.6 R-Package for survival analysis

Survival analysis in R Programming Language deals with the prediction of events at a specified time. It deals with the occurrence of an interesting event within a specified time and failure of it produces censored observations i.e incomplete observations.

The R package named **SURVIVAL** is used to carry out survival analysis. This package contains two functions namely **Surv()** and **survfit()**. Surv() which takes the input data as a R formula and creates a survival object among the chosen variables for analysis. Then we use the function survfit() to create a plot for the analysis. The syntax for creating survival analysis in R-software is:

Surv(time,event)

survfit(formula)

Following is the description of the parameters used-

- time is the follow up time until the event occurs.
- event indicates the status of occurrence of the expected event.
- formula is the relationship between the predictor variables.

6.7 Survival Analysis of Leukemia data

Let's look at the following data set in the survival library in R. The following are times to relapse (weeks) for 21 leukemia patients receiving control treatment data give by **Laud Randy Amofah** March 2020.

$$1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23$$

Structure of data:

- Variable under study: Time in weeks
- $n = 21$

Descriptive statistics of data

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   4.000   8.000   8.667  12.000  23.000
```

The following is the R-code.

R-code to fit the Leukemia data for Weibull distribution

```
library(survival)
t=c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23)
S1=Surv(t)
s1=survreg(Surv(t)~1,S1, dist="weibull",scale=0)
summary(s1)
plot(S1,xlab="Time",ylab="S(t)"),lty=1,col=6,main= "Non-parametric
estimator of survival function")
legend(x="topright",legend=c("Survival function for Leukemia
data","Confidence interval for Leukemia data"),col=6,lty=1:2)
```

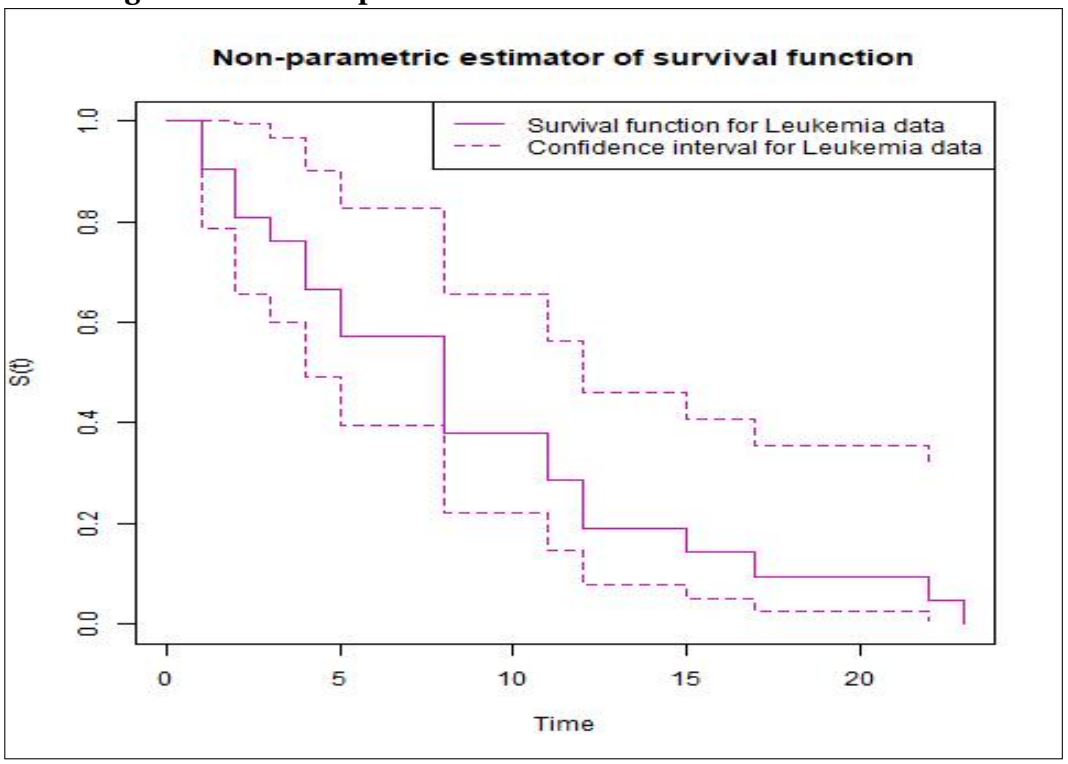
The following is the R-output obtained using survival package.

R-output of fitting Weibull distribution to Leukemia patient data

```
Call: survreg(formula = Surv(t) ~ 1, data = S1,
        dist = "weibull", scale = 0)
             Value Std. Error     z      p
(Intercept)  2.249      0.168 13.40 <2e-16
Log(scale)  -0.315      0.174 -1.82  0.069

Scale= 0.73
Weibull distribution
Loglik(model)= -64.9   Loglik(intercept only)= -64.9
Number of Newton-Raphson Iterations: 6
n= 21
```

Figure 6.4 Survival plot of Weibull distribution for Lukemia data



6.8 Survival Analysis of Ovarian data

Ovarian dataset comprises a cohort of ovarian cancer patients and respective clinical information, including the time patients were tracked until they either died or were lost to follow-up (futime), whether patients were censored or not (fustat).

Structure of data:

- Variable under study: Time
- $n = 26$

Descriptive statistics of data

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   59.0   368.0   476.0   599.5   794.8  1227.0
```

Table 6.1 Ovarian Data

Sr. No.	**futime**	**fustat**
1	59	1
2	115	1
3	156	1
4	421	0
5	431	1
6	448	0
7	464	1
8	475	1
9	477	0
10	563	1
11	638	1
12	744	0
13	769	0
14	770	0
15	803	0
16	855	0
17	1040	0
18	1106	0
19	1129	0
20	1206	0
21	1227	0
22	268	1
23	329	1
24	353	1
25	365	1
26	377	0

The R-code for fitting the distribution is as follows.

R-code to fit the Overian data for Weibull distribution

```
library(survival)
t= Surv(time = ovarian$futime, event = ovarian$fustat)
fit_1= survfit(t~ 1,data = ovarian)
summary(fit_1)
plot(t, conf.int=T, xlab="Time", ylab= "Survival Function", main= "Non-parametric estimator of survival function",cex=.6, col="red")
legend(x="bottomright",legend=c("Survival function of Overian data"),col=c("red"),lty=1)
```

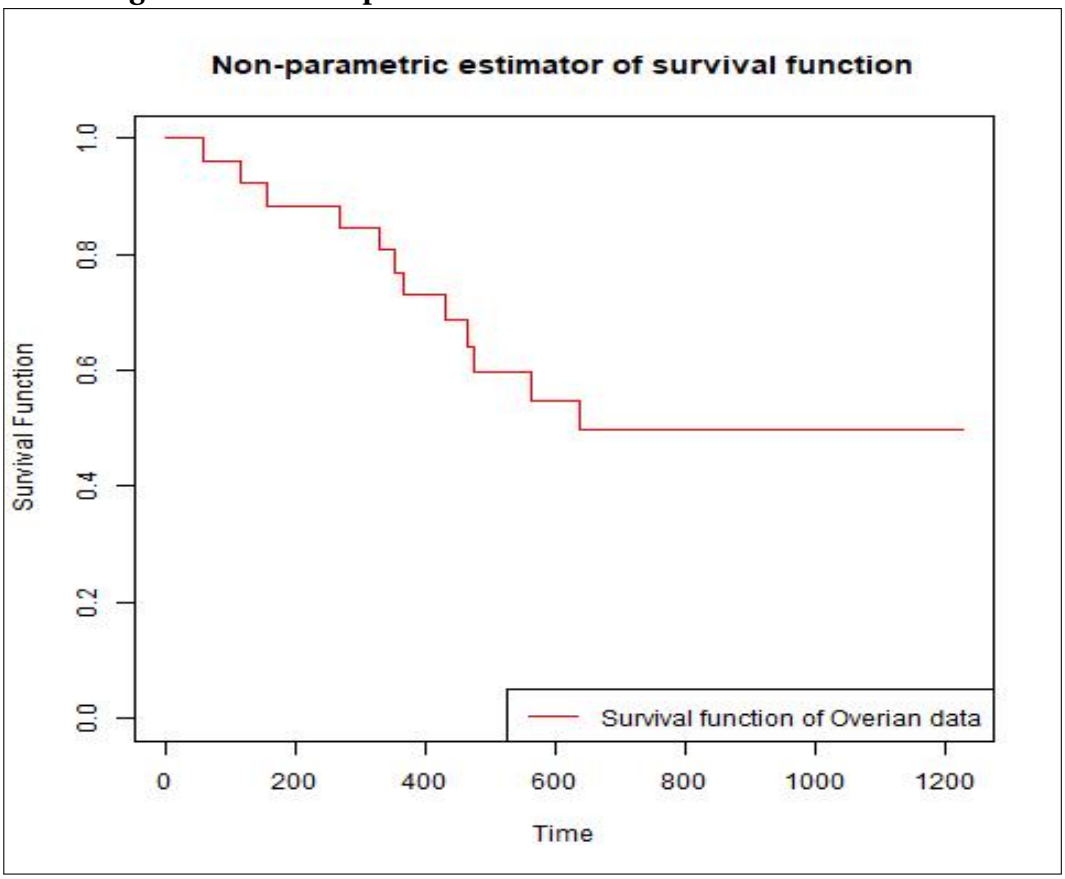
The following is the output of fitting Weibull distribution to Overian data obtained from R-software using Survival package.

R-output of fitting Weibull distribution to Overian data

```
Call: survfit(formula = t ~ 1, data = ovarian)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   59     26       1    0.962  0.0377        0.890        1.000
  115     25       1    0.923  0.0523        0.826        1.000
  156     24       1    0.885  0.0627        0.770        1.000
  268     23       1    0.846  0.0708        0.718        0.997
  329     22       1    0.808  0.0773        0.670        0.974
  353     21       1    0.769  0.0826        0.623        0.949
  365     20       1    0.731  0.0870        0.579        0.923
  431     17       1    0.688  0.0919        0.529        0.894
  464     15       1    0.642  0.0965        0.478        0.862
  475     14       1    0.596  0.0999        0.429        0.828
  563     12       1    0.546  0.1032        0.377        0.791
  638     11       1    0.497  0.1051        0.328        0.752
```

Figure 6.5 Survival plot of Weibull distribution for Overian data



References

- C. D. Lai, D. N. Murthy, and M. Xie, "Weibull Distributions and Their Applications" Springer Handbooks, Berlin, Germany, 2006.
- Mohammad A. Al-Fawzan, "Methods for Estimating the Parameters of the Weibull Distribution", Riyadh 11442, Saudi Arabia, May 2000.
- Antonio Sanhueza, Víctor Leiva, Narayanaswamy Balakrishnan, "The Generalized Birnbaum–Saunders Distribution and Its Theory, Methodology, and Application", McMaster University, Hamilton, Ontario, Canada, June 2015.
- Vito Ricci, "FITTING DISTRIBUTIONS WITH R", February 2005.
- Applied Mathematical Sciences, Vol. 8, 2014, no. 83, 4137 - 4149 HIKARI Ltd, http://dx.doi.org/10.12988/ams.2014.4538
- Mathematical Problems in Engineering Volume 2021 https://doi.org/10.1155/2021/9175170
- https://www.weibull.com/hotwire/issue14/relbasics14.htm
- https://www.tutorialspoint.com/r/r_survival_analysis.htm
- https://www.itl.nist.gov/div898/handbook/eda/section3/eda3668.htm