Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2023, 2023:44
https://doi.org/10.28919/cmbn/7937
ISSN: 2052-2541

PARAMETER ESTIMATION OF WEIBULL DISTRIBUTION ON TYPE III CENSORED SURVIVAL DATA BY MAXIMUM LIKELIHOOD ESTIMATOR METHOD: CASE STUDY OF LUNG CANCER PATIENT DATA AT DR. KARIADI HOSPITAL SEMARANG

ARDI KURNIAWAN*, RENDI KURNIA RAHMANTA, EKO TJAHJONO

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia

Abstract: Research on survival data is quite important. Many modeling problems can be made related to survival data. One of the survival data methods is the Type III Censored method, which is a type of censoring that is limited by time, with individuals entering the study at different times during a certain period of time. In this study, the research focus is to obtain point estimates of the Weibull distribution parameters using the Maximum Likelihood Estimator method. The method is used to examine the survival data of Weibull distribution censored type III on secondary data of patients with lung cancer at Dr. Kariadi Hospital Semarang. Based on the results of the analysis, the Weibull distribution with a point estimator for the scale parameter is $\hat{\alpha} = 94.875997$ and the shape parameter is $\hat{\beta} = 1.638958$.

Keywords: Weibull distribution; type III censored; maximum likelihood.

2020 AMS Subject Classification: 62F40.

1. INTRODUCTION

Survival data analysis is an analysis that discusses the survival of an object or individual in certain operational conditions. This analysis is usually applied to observations of the durability of a product to observations in the field of human disease that can be used to improve product quality

*Corresponding author
E-mail address: ardi-k@fst.unair.ac.id
Received March 03, 2023

and develop modern treatment methods [1]. Survival analysis is an analysis of survival that is usually used to analyze the survival of patients against a disease [5].

In observing survival data, various distribution models can be used. One of the distributions that can be used is the Weibull distribution [5]. Weibull distribution is one of the continuous random variable distributions that is often used to analyze the reliability value of a system. This distribution is famous for its flexible distribution. One of its flexibility can be seen in changing this distribution into other distributions, such as the exponential distribution, depending on changes in the scale and shape parameters.

In survival data analysis research, survival time data can be in the form of censored and uncensored data. Censored observations occur if the survival data of the observed individuals are not known with certainty. There are three types of censoring in survival data observation: type I, type II, and type III. Type I censored is when the test is stopped after a specific time, type II censored is when the test is stopped after a certain number of failures are obtained. In contrast, type III censored is an observation that is limited by the time of each patient entering the observation at different times. This type of censoring is often used in research on the survival of patients with a disease because some patients rarely enter the observation simultaneously.

In survival data analysis, research is often carried out on parameter estimation of the distribution and type of survival time data which aims to infer statistical values. One of them is in the thesis [3] on "Confidence Interval of Rayleigh Distribution Parameters on Interval Censored Survival Data with Bootstrap Method". So far, statistical inference in classical statistics is generally based on the normal or approximate normal distribution. In 1979 statistician Efron introduced the Bootstrap method to be utilized in statistical inference. In previous studies, model estimation using a program has also been carried out in research [4] on the role of the S-Plus 2000 program in nonparametric regression models with the Faorier series approach to rainfall model estimation.

Efron first introduced the Bootstrap method in 1979. The name Bootstrap is taken from the phrase "to pull oneself up by one's bootstraps," which means standing on one's own feet. This bootstrap approach uses a resampling method. Based on the characteristics of the Bootstrap method, which is a data-based simulation method for certain more straightforward statistical inferential purposes with the help of modern computer power [2]. Several studies related to the utilization of the bootstrap method in the field of life test analysis have been conducted by Akhmad Fauzy (2014), with the focus of his research being the use of the Bootstrap method as an alternative method for predicting confidence intervals for parameters (case study: single, double and multiple type-II censored exponentially distributed data).

Based on the description above, the author wants to know the confidence interval results for the type III censored Weibull distribution parameters with the Bootstrap method. In this study, it does not require the possibility of the Weibull distribution parameter estimation results obtained in the form of implicit functions. Therefore the numerical iteration method is used if this happens.

2. PRELIMINARIES

A. Censored Data

Censored data is data where the value of the dependent variable under study does not provide complete information or cannot be fully observed. This is due to the limitations of the observations made or the observed individuals leaving the study. Censored data contains only partial information about the random variable of interest but affects statistical notions and calculations. Several kinds of methods are often used in life test experiments, one of which is Type III Censored. Censored type III is said if individuals or objects enter the study at different times during a certain period of time.

Likelihood function on type III censored data from observation data, (t_i, δ_i), $i = 1, 2, \ldots, n$ obtained:

$$L(\theta) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{1}$$

δ_i is a censoring indicator, worth one if the data is not censored and worth zero if the data is censored. t_i is obtained from $\min(T_i, C_i), i = 1, 2, \ldots, n$. T_i is the individual's lifetime and C_i is the censoring time [5], where in this study, the value of censored time is determined equally for all n of the research individuals according to the time limit of the research conducted.

B. Weibull Distribution

According to [4], the Weibull distribution is a distribution that describes extreme events such as the lifetime of living things. The Weibull distribution is most widely used in lifetime distribution models. Suppose a continuous random variable T is Weibull distributed, with α ss the scale parameter and β as the shape parameter, then the PDF is:

$$f(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], for\ x > 0 \\ 0, for\ another \end{cases} \tag{2}$$

While the CDF of the Weibull distribution is

$$F(x) = 1 - exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], x > 0 \tag{3}$$

The survival function of the Weibull distribution is

$$S(x) = \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], t > 0 \tag{4}$$

C. Point Estimation and Maximum Likelihood Estimator (MLE)

1. Point Estimation

A sample distribution from a population is useful for making inferences about the population. Two important issues in statistical inference are estimation and hypothesis testing. One type of estimation is point estimation. According to Graybill et al. (1963), if there are values of some statistics $t(Y_1, Y_2, Y_3, \ldots, Y_n)$ that represent or estimate the unknown parameter θ, then each statistic $t(Y_1, Y_2, Y_3, \ldots, Y_n)$ is called a point estimator.

2. Maximum Likelihood Estimator (MLE)

The following will explain some of the defines of Hogg and Craig (1995) :

Definition 2.1: Let be $Y_1, Y_2, Y_3, \ldots, Y_n$ an iid random variable from a distribution with PDF $f(y_i, \theta)$, for $\theta \in \Omega$, where Ω is the set of parameters. The joint PDF between $f(y_1;\theta) f(y_2;\theta) \ldots f(y_n;\theta)$. If the joint PDF is expressed as a function of θ then it is called the Likelihood function which is denoted L or written as follows:

$$L(\theta; y_1, y_2, \ldots, y_n) = f(y_1;\theta) f(y_2;\theta) \ldots f(y_n;\theta) \tag{5}$$

Definition 2.2: If the statistic $\hat{\theta} = t(Y_1, Y_2, Y_3, \ldots, Y_n)$ maximizes the Likelihood function $L(\theta; y_1, y_2, \ldots, y_n); \theta \in \Omega$, then the statistic $\hat{\theta} = t(Y_1, Y_2, Y_3, \ldots, Y_n)$ is called the Maximum Likelihood Estimator (MLE) of θ.

3. METHODOLOGY

A. Data

The data source used in this study is secondary data obtained from research [6] related to data on patients with cancer at Dr. Kariadi Hospital Semarang.

B. Analysis Procedure

The research method related to the purpose of this paper is as follows.

1. Assume that $y_1, y_2, \dots, y_n$ are survival data from a random sample of size n derived from the Weibull (α, β) distribution.
2. Determine the shape of the Weibull distribution likelihood function on type III censored survival data.
3. Determine the ln-likelihood function of the Weibull distribution on type III censored survival data.
4. Differentialize the ln-likelihood function with respect to the Weibull distribution parameters.
5. The result of the differential is equated to zero and solved.

4. MAIN RESULTS

A. Determining the shape of the Weibull Distribution parameter point estimate with the Maximum Likelihood Estimator Method

Point estimation of Weibull distribution parameters on type III censored survival data with the Maximum Likelihood Estimator Method used the following steps:

1. Suppose $y_1, y_2, \dots, y_n$ is a positive random variable in type III mutually iid censored survival data from the Weibull distribution with parameters α and β, so it can be written as follows:

$$y_i \sim Weibull\ (\alpha, \beta), \text{with } i = 1, \dots, \text{n} \tag{6}$$

The failure time of the i-th observation in type III censored survival data is expressed as y_i.

2. Based on the shape of the Weibull distribution survival function and the shape of type III censored likelihood function, the following type III Weibull-distributed censored likelihood function is obtained:

$$f(\mathrm{y}_i) = \frac{\beta}{\alpha}\left(\frac{y_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]$$

$$S(\mathrm{y}_i) = \exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]$$

So: $$L(\alpha,\beta;\mathrm{y}_1,\mathrm{y}_2,...,\mathrm{y}_n)=\prod_{i=1}^{n}\left\{\left[\frac{\beta}{\alpha}\left(\frac{y_i}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]^{\delta_i}\left[\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]^{1-\delta_i}\right]\right\} \quad (7)$$

3. The next step is determining the ln-likelihood function of type III censored Weibull distribution. From the likelihood function equation, the ln-likelihood function is obtained as follows:

$$\mathrm{Ln}\,\ell L(\alpha,\beta;\mathrm{y}_1,\mathrm{y}_2,...,\mathrm{y}_n)=\ln\prod_{i=1}^{n}\left\{\left[\frac{\beta}{\alpha}\left(\frac{y_i}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]^{\delta_i}\left[\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]^{1-\delta_i}\right]\right\}$$

$$=\sum_{i=1}^{n}\ln\left\{\left[\left(\frac{\beta}{\alpha}\right)\left(\frac{y_i}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]^{\delta_i}\left[\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]^{1-\delta_i}\right\}$$

$$=\sum_{i=1}^{n}\delta_i\left[\ln\left(\frac{\beta}{\alpha}\right)+\ln\left(\frac{y_i}{\alpha}\right)^{\beta-1}+\ln\left[\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]\right]+\sum_{i=1}^{n}(1-\delta_i)\ln\left[\exp\left[-\left(\frac{y_i}{\alpha}\right)^{\beta}\right]\right]$$

$$=\sum_{i=1}^{n}\delta_i\ln\beta+\sum_{i=1}^{n}\delta_i\beta\ln y_i-\sum_{i=1}^{n}\delta_i\ln y_i-\sum_{i=1}^{n}\delta_i\beta\ln\alpha-\sum_{i=1}^{n}\left(\frac{y_i}{\alpha}\right)^{\beta} \quad (8)$$

4. After obtaining the ln-likelihood function in equation (8), the next step is to derive the ln-likelihood function for the Weibull (α,β) distribution parameters, namely:

   The ln-likelihood derivative of the parameters α nd equalized to zero is as follows:

$$0=\frac{\partial(\ln L(\alpha,\beta;\mathrm{y}_1,\mathrm{y}_2,...,\mathrm{y}_n))}{\partial\alpha}=\frac{\partial\left(\sum_{i=1}^{n}\delta_i\ln\beta+\sum_{i=1}^{n}\delta_i\beta\ln y_i-\sum_{i=1}^{n}\delta_i\ln y_i-\sum_{i=1}^{n}\delta_i\beta\ln\alpha-\sum_{i=1}^{n}\left(\frac{y_i}{\alpha}\right)^{\beta}\right)}{\partial\alpha}$$

So: $$0=-\frac{\sum_{i=1}^{n}\delta_i\beta}{\alpha}-\sum_{i=1}^{n}\left(y_i\right)^{\beta}(-\beta\alpha^{(-\beta-1)})$$

From the results obtained:

$$\hat{\alpha}=\left[\frac{\sum_{i=1}^{n}\left(y_i\right)^{\beta}}{\sum_{i=1}^{n}\delta_i}\right]^{\frac{1}{\beta}} \quad (9)$$

The ln-Likelihood derivative of the parameter β is as follows:

$$0=\frac{\partial(\ln L(\alpha,\beta;\mathrm{y}_1,\mathrm{y}_2,...,\mathrm{y}_n))}{\partial\beta}=\frac{\partial\left(\sum_{i=1}^{n}\delta_i\ln\beta+\sum_{i=1}^{n}\delta_i\beta\ln y_i-\sum_{i=1}^{n}\delta_i\ln y_i-\sum_{i=1}^{n}\delta_i\beta\ln\alpha-\sum_{i=1}^{n}\left(\frac{y_i}{\alpha}\right)^{\beta}\right)}{\partial\beta}$$

$$=\frac{\sum_{i=1}^{n}\delta_i}{\beta}+\sum_{i=1}^{n}\delta_i(\ln y_i)-\sum_{i=1}^{n}\delta_i(\ln\alpha)-\left(\frac{1}{\alpha^{\beta}}\right)\sum_{i=1}^{n}(y_i)^{\beta}\ln\mathrm{y}_i+\left(\frac{1}{\alpha^{\beta}}\right)\sum_{i=1}^{n}(y_i)^{\beta}\ln\alpha$$

By substituting the value of the parameter point estimate result $\alpha=\left[\frac{\sum_{i=1}^{n}(y_i)^{\beta}}{\sum_{i=1}^{n}\delta_i}\right]^{\frac{1}{\beta}}$ to the equation and equated to zero, we get the following:

$$0=\frac{\sum_{i=1}^{n}\delta_i}{\beta}+\sum_{i=1}^{n}\delta_i(\ln y_i)-\sum_{i=1}^{n}\delta_i\ln\left[\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}(y_i)^{\beta}}\right]^{-\frac{1}{\beta}}-\left(\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}(y_i)^{\beta}}\right)\sum_{i=1}^{n}(y_i)^{\beta}\ln\mathrm{y}_i+\left(\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}(y_i)^{\beta}}\right)\sum_{i=1}^{n}(y_i)^{\beta}\ln\left[\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}(y_i)^{\beta}}\right]^{-\frac{1}{\beta}}$$

The final result of the parsing is as follows.

$$\frac{\sum_{i=1}^{n}\delta_i\sum_{i=1}^{n}(y_i)^{\hat{\beta}}\ln\mathrm{y}_i}{\sum_{i=1}^{n}(y_i)^{\hat{\beta}}}-\frac{\sum_{i=1}^{n}\delta_i}{\hat{\beta}}-\sum_{i=1}^{n}\delta_i(\ln y_i)=0 \tag{10}$$

With the least squares method approach, $\hat{\beta}$ is obtained:

$$\hat{\beta}=\frac{n\sum_{i=1}^{n}\left(\ln\left[\ln\left[\frac{1}{1-F(y_i)}\right]\right]\right)^2-\left(\sum_{i=1}^{n}\left(\ln\left[\ln\left[\frac{1}{1-F(y_i)}\right]\right]\right)\right)^2}{n\sum_{i=1}^{n}\ln y_i\left(\ln\left[\ln\left[\frac{1}{1-F(y_i)}\right]\right]\right)-\sum_{i=1}^{n}\ln\left[\ln\left[\frac{1}{1-F(y_i)}\right]\right]\sum_{i=1}^{n}\ln y_i}$$

B. Application to Data on Lung Cancer Patients Who Get Treatment at Dr. Kariadi Hospital Semarang

In the discussion of this research, secondary data is obtained from research [6]. The data is the survival time of lung cancer patients who get treatment at Dr. Kariadi Hospital Semarang. Suppose y_i is data stating the failure/death time of 50 patients with lung cancer. Furthermore, testing is carried out to determine whether the survival data of lung cancer patients who get treatment at Dr. Kariadi Semarang Hospital is Weibull distributed at the 5% significance level. The hypothesis used:

H_0 : type III censored data of survival time of lung cancer patients with Weibull distribution

H_1 : type III censored data of survival time of lung cancer patients is not Weibull distributed

Based on the output of the Kolmogorov-Smirnov test and the Chi-Square test, the conclusion is not to reject H_0 or in other words, accept H_0.

The p-value of type III censored data (y_i) of lung cancer patients who received treatment at Dr. Kariadi Hospital Semarang was obtained as 0.18419 in the Kolmogorov-Smirnov test and the Chi-Square test 0.11864. Because the p-value obtained is greater than the significance level, the censored data type III survival time of lung cancer patients who received treatment at Dr. Kariadi Hospital Semarang is Weibull distributed.

The next step will be to find the value of the point estimator using the Maximum Likelihood Estimator method of the Weibull distribution of type III censored survival data on lung cancer patients who received treatment at Dr. Kariadi Hospital Semarang with the help of R- 3.1.2 software.

Based on the calculation, in this case, using the program, the point estimate value $\hat{\alpha} = 94.78144$ and $\hat{\beta} = 1.638958$.

Based on equation (4) the estimated survival function equation, namely the probability of lung cancer patients at Dr. Kariadi Hospital Semarang being able to survive beyond time y, can be obtained as follows:

$$\hat{S}(y) = 1 - F(y)$$

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

$$= exp\left[-\left(\frac{y}{94.78144}\right)^{1.638958}\right]$$

For example, suppose it is desired to calculate the expected survival function at y = 5, it will be obtained:

$$\hat{S}(5) = exp\left[-\left(\frac{y}{\alpha}\right)^{\beta}\right]$$

$$= exp\left[-\left(\frac{5}{94.78144}\right)^{1.638958}\right]$$

$$= 0.9919819$$

This means the probability of lung cancer patients to survive more than 5 days is 99.2%.

5. CONCLUSION

Based on the analysis results, the value of the point estimator $\hat{\alpha}$ is obtained as follows.

$$\hat{\alpha} = \left[\frac{\sum_{i=1}^{n}(y_i)^{\beta}}{\sum_{i=1}^{n}\delta_i}\right]^{\frac{1}{\beta}}$$

Then for the application to the data of patients with lung cancer at Dr. Kariadi Hospital Semarang resulted in $\hat{\alpha} = 94.875997$ and $\hat{\beta} = 1.638958$.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] R.E. Barlow, F. Proschan, L.C. Hunter, Mathematical theory of reliability, SIAM, Philadelphia, 1996.

[2] E. Elviana, J. Purwadi, Parameters estimation of Rayleigh distribution in survival analysis on type II censored data using the Bayesian method, J. Phys.: Conf. Ser. 1503 (2020), 012004. https://doi.org/10.1088/1742-6596/1503/1/012004.

[3] M. Krisdiyanti, Parameter estimation of Rayleigh distributed type III censored data by maximum likelihood method, Thesis, Jendral Soedirman University. (2021).

[4] J.F. Lawless, Statistical models and methods for life time data, John Wiley & Sons, New York. (1982).

[5] P.J. Smith, Analysis of failure and survival data, Chapman and Hall/CRC, New York. (2002).

[6] D.T. Utami, Analysis of life test data of lung cancer patients at Dr. Kariadi Hospital Semarang. Thesis, FMIPA. State University of Semarang. (2015).