Confidence Intervals Estimation for Survival Function in Log-Logistic Distribution and Proportional Odds Regression Based on Censored Survival Time Data

Kamil ALAKUŞ*, Necati Alp ERİLLİ

Ondokuz Mayıs University, Faculty of science and Arts, Department of Statistics, 55139 Turkey

Abstract

Log-logistic and Weibull distributions have both accelerated survival time property. The log-logistic distribution has also proportional odds property. Log-logistic distribution has unimodal hazard curve which changes direction. Link [6,7] presented a confidence interval estimate of survival function using Cox's proportional hazard model with covariates. Her idea more recently extended by [1] to the exponential distribution and [2] to exponential proportional hazard model, respectively. The same idea has been extended to the Weibull proportional hazard regression model by [3]. In this study, it is formed on confidence interval for log-logistic distribution survival function for any values of the time provided that the survival times have a log-logistic distributed random variable. It is also extended the same results to the proportional odds regression. A Real time data and a simulation data examples are also considered in the study for illustration the discussed confidence interval.

Keywords: Confidence interval; Hazard function; Point estimation; Survival analysis; Survival function; Log-logistic distribution; Proportional odds regression

Introduction

There are two types of estimation for any identity. One is point estimation and the other is confidence interval estimation. In Survival analysis literature, confidence interval estimate for the survival function is not new. Especially confidence interval estimate for the baseline survival function is extensively studied with many authors. For example, for Kaplan-Meier survival function confidence interval estimate is studied using Greenwood formulae by [4,5] and many others. In Cox's proportional hazard model, [6,7] formed log transformed confidence interval for survival function with covariates. Her idea more recently extended by [1] to the exponential distribution and [2] to exponential proportional hazard model, respectively. Weibull proportional hazard function is also investigated by [3]. Interval estimate for survival function is generally useful in the analysis of survival or life time data. In this study, symmetric type and proportional odds transformed confidence interval approach is developed for log-logistic survival function without covariates and proportional odds regression survival function with covariates.

The plan of this study is given as follows. In the next section, the log-logistic distribution and proportional odds regression and their important functions are presented. In section 3, it is formed confidence interval estimate for the survival function of log-logistic distribution and proportional odds regression model too. In section 4; it has given a real data example and as an extension of the real data, simulation study for illustrating the proposed method in this study. The study completed with a discussion section.

Log-logistic distribution and proportional odds regression model

Log logistic distribution: The log-logistic distribution has the proportional odds property and the distribution is the natural one to use in conjunction with the proportional odds model. Cox and Oakes [8] demonstrated that the log-logistic distribution is the only one that shares the accelerated survival time property and proportional odds property. Situations in which the hazard function changes, direction

of hazard function can arise. For example, a patient faces an increasing hazard of death in heart transplantation over the first few days or after the transplant, while the body adapts to the new recovers. The hazard then decreases with time as the patient recovers. In situations such as this, a unimodal hazard function may be appropriate.

A particular form of unimodal hazard is the function

$$\lambda(t) = \left(\alpha \lambda^{\alpha} t^{(\alpha-1)}\right)\left[1 + (\lambda t)^{\alpha}\right]^{-1} \qquad (1)$$

This hazard function decreases monotonically if $\alpha \leq 1$. If $\alpha > 1$, than the hazard has a stable mode. The survival function corresponding to the hazard function in equation (1) is given below

$$S(t) = \exp\left(-\int_0^t \left(\alpha \lambda^{\alpha} s^{(\alpha-1)}\right)\left[1 + (\lambda s)^{\alpha}\right]^{-1} ds\right) = \left[1 + (\lambda t)^{\alpha}\right]^{-1} \quad (2)$$

Probability density function is also given below

$$f(t) = \lambda(t)S(t) = \left(\alpha \lambda^{\alpha} t^{(\alpha-1)}\right)\left[1 + (\lambda t)^{\alpha}\right]^{-2}, t > 0 \qquad (3)$$

This is the density function of a random variable T which has a log-logistic distribution with parameters α and λ. The distribution is so called because the variable logT has log-logistic distribution and a symmetric distribution whose probability density function is very similar to the normal distribution [9].

The suitability of the log-logistic distribution for the analysis of a data set can be empirically checked using a linear relationship derived

*Corresponding author: Kamil ALAKUŞ, Ondokuz Mayıs University, Faculty of science and Arts, Department of Statistics, 55139 Turkey, E-mail: kamilal@omu.edu.tr

from expression of S(t) and $F(t)=1-S(t)$. The odds of surviving beyond time t are $S(t)/F(t)=(\lambda t)^{-\alpha}$ and consequently log odds of survival beyond t can be expressed as $\log\left\{\dfrac{S(t)}{F(t)}\right\} = -\alpha\log\lambda - \alpha\log t$ or equivalently

$$\log\left\{\frac{F(t)}{S(t)}\right\} = \alpha\log\lambda + \alpha\log t \qquad (4)$$

That is, the log-logistic distribution corresponds to a linear model for the log odds of failure over the logarithm of time, with slope α [10,11]. The Kaplan-Meier sample estimate $\hat{S}_{KM}(t)$ can be used to calculate the log odds and a plot of them versus log t should follow approximately a straight line for the log-logistic model in order to be suitable.

Proportional odds regression model

The application of accelerated survival time and proportional odds models to the analysis of reliability data has recently been described by [12]. The general proportional odds model for survival data was introduced by [13,14] describes the proportional odds model.

The log-logistic distribution is commonly extended to include a vector of covariates x by reformulating the survival function is given as follows

$$S(t;x) = \left[1 + (\lambda t)^{\alpha}\exp\left(\alpha\theta^{T}x\right)\right]^{-1} \qquad (5)$$

Under the accelerated failure time model, the hazard of death at the time of t is given below

$$h(t;x) = \left(\alpha\lambda^{\alpha}t^{(\alpha-1)}\right)\left[1 + (\lambda t)^{\alpha}\exp\left(\alpha\theta^{T}x\right)\right]^{-1} \qquad (6)$$

Thus the survival time for the ith observation also has a log-logistic distribution and therefore it has both accelerated failure time and proportional odds property.

Confidence intervals for survival function

Log-logistic distribution: First multiply, both numerator and denominator, by α with equation (2). The equation value will not change. Therefore we can re-write equation (2) as follows

$$S(t_i) = \alpha\left[\alpha + \alpha(\lambda t_i)^{\alpha}\right]^{-1} \qquad (7)$$

If we take natural logarithm of $\alpha(\lambda)^{\alpha}$ identity, we can get the result as $\log\alpha + \alpha\log\lambda$. Now let $R_i = \log\alpha + \log\lambda$ be the score (risk score) function. Then we denote $\log\lambda$ with β_0 and $\log\alpha$ with β_1, the score function can be written as $R_i = \beta_0 + \beta_1$. Therefore the survival function given (7) can be re-written as follows

$$S(t_i) = \alpha\left[\alpha + e^{R_i}t_i^{\alpha}\right]^{-1} \qquad (8)$$

Estimated survival function is also given by

$$\hat{S}(t_i) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_i}t_i^{\hat{\alpha}}\right]^{-1} \qquad (9)$$

As a result we can prepare for confidence intervals of survival function for the log-logistic distribution. To do this, we first can form confidence intervals for R_i then extended it to the survival function. For the score function R_i 100(1-α)% confidence interval is given by

$$\Pr\left\{\hat{R}_i - z_{\alpha/2}se\left(\hat{R}_i\right) \le R_i \le \hat{R}_i + z_{\alpha/2}se\left(\hat{R}_i\right)\right\} = 1-\alpha \qquad (10)$$

or

$$\Pr\left(\hat{R}_{\text{low}(i)} \le R_i \le \hat{R}_{\text{upp}(i)}\right) = 1-\alpha \qquad (11)$$

Here $Z\alpha/2$ denotes coordinate value of standard normal distribution at the significance level of $\alpha/2$ and $se(\hat{R}i)$ is also denotes standard error of estimated score function. The estimated standard error of the score function is calculated using by

$$se\left(\hat{R}_i\right) = \left\{x_i^{T}Var\left(\hat{a}\right)x_i\right\}^{1/2} \qquad (12)$$

where, $x_i^{T} = \begin{bmatrix}1 & 1\end{bmatrix}$ is a unit column vector in this simple model and $Var(\hat{a})$ is also variance-covariance matrix of estimated parameters.

We can easily form a 100(1-α)% confidence intervals of survival function using score function confidence intervals. So that, the confidence intervals for survival function of a log-logistic distribution are given by $\hat{S}_{\text{low}}(t_i) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_{UPP(i)}}t_i^{\hat{\alpha}}\right]^{-1}$ for lower limit and $\hat{S}_{\text{upp}}(t_i) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_{LOW(i)}}t_i^{\hat{\alpha}}\right]^{-1}$ for upper limit, respectively. So, 100(1-α)% confidence intervals for survival function of the log-logistic distribution can given as following

$$\Pr\left\{\hat{S}_{\text{low}}(t_i) \le S(t_i) \le \hat{S}_{\text{upp}}(t_i)\right\} = 1-\alpha \qquad (13)$$

Proportional odds regression model

Let $R_i = \log\alpha + \log\lambda + \alpha\theta^T x_i$ be score function for i-th observation. If we denote $\alpha\log\lambda$ with β_0 and $\log\alpha$ with β_1, then the score function for i-th observation can be written as $R_i = \beta_0 + \beta_1 + \beta^T x_i$. Here β^T is ($p\times1$) column vector and equals to $\alpha\theta^T$.

Let $\beta_0 + \beta_1 + \beta^T$ denote as $\bar{a}^T$ a ($k\times1$) size column vector and y_i also a same size column vector too. Then the score function for i-th observation can be written as $R_i = \bar{a}^T y_i$. So that, the survival function given in (2) can be written as in equation (14)

$$S(t_i;y_i) = \alpha\left[\alpha + e^{R_i}t_i^{\alpha}\right]^{-1} \qquad (14)$$

Estimated survival function is also given by equation (15)

$$\hat{S}(t_i;y_i) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_i}t_i^{\hat{\alpha}}\right]^{-1} \qquad (15)$$

Therefore we can prepare confidence intervals of survival function for the proportional odds regression. As we made for log-logistic distribution, firstly we can form confidence intervals for R_i then extended it to the survival function. For the score function R_i 100(1-α)% confidence interval is given by equation (10) or equation (11). The estimated standard error of the score function for i-th observation is calculated using equation (16)

$$se\left(\hat{R}_i\right) = \left\{\hat{y}_i^{T}Var\left(\hat{\bar{a}}\right)_i\right\}^{1/2} \qquad (16)$$

where, $y_i^{T} = \begin{bmatrix}1 & 1 & x_{1i} & \cdots & x_{ki}\end{bmatrix}$ a column vector in this model and $Var(\hat{\bar{a}})$ is also variance-covariance matrix of estimated parameters. In this model, the estimated variance-covariance matrix might be given as follows

$$Var\left(\hat{\bar{a}}\right) = \begin{bmatrix} Var(\hat{\beta}_0) & \cdots Cov(\hat{\beta}_0,\hat{\beta}_k) \\ \vdots & \vdots & \vdots \\ Cov(\hat{\beta}_0,\hat{\beta}_k) & \cdots & Var(\hat{\beta}_k) \end{bmatrix} \qquad (17)$$

A 100(1-α)% confidence intervals of survival function uses

confidence intervals of score function. Namely, the confidence intervals for survival function of a proportional odds model are given by $\hat{S}_{low}\left(t_i; y_i\right) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_{UPP(i)}} t_i^{\hat{\alpha}}\right]^{-1}$ for lower limit and $\hat{S}_{upp}\left(t_i; y_i\right) = \hat{\alpha}\left[\hat{\alpha} + e^{\hat{R}_{LOW(i)}} t_i^{\hat{\alpha}}\right]^{-1}$ for upper limit, respectively. So, $100(1-\alpha)\%$ confidence intervals for survival function of the proportional odds regression can be shown as follows

$$\Pr\left\{\hat{S}_{low}\left(t_i; y_i\right) \le S\left(t_i; y_i\right) \le \hat{S}_{upp}\left(t_i; y_i\right)\right\} = 1-\alpha \qquad (18)$$

Model selection criterion

In this study we tested performance of proposed model with log-likelihood value and besides with AIC and BIC model selection criterions. The Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. It was developed by [15]. Bayesian Information Criterion (BIC) is a criterion for model selection among a class of parametric models with different numbers of parameters. It has been introduced by [16].The AIC and BIC is given as follows, respectively

$$AIC = -2 \times loglik + 2 \times p \qquad (19)$$

$$BIC = -2 \times loglik + p \times \ln(n)s \qquad (20)$$

When data has small samples size, corrected AIC can be given

$$AICc = -2 \times \log lik + \frac{2 \times p \times n}{(n - p - 1)} \qquad (21)$$

where p is defined as number of free parameters.

Given a set of candidate models for the data, the preferred model is the one with the smallest value of Log Likelihood, AIC and BIC.

Illustrative example

In this section a real data illustration considered to confidence intervals for survival function which we discussed it earlier sections in this study. For this reason, we first give some information about the data just in the next subsection. Second, we use the data for illustrating confidence intervals estimation for the survival function of the proportional odds regression model.

Real time data: Ovarian cancer study

Data are from [17] on ovarian cancer. The data are taken from [18]. The ovarian cancer frame includes the survival times (in days) and indicator variable (status) of death or censoring plus the following 4 additional variables on each patient. These are patient's age (age), an indicator of the extent of the residual disease (residual.dz), treatment given (rx) and measure of performance score or functional status using the Eastern Cooperative Oncology Group's scale (ecog.ps). The survival analysis chapter in the S-Plus documentation describes these data sets further and illustrates survival analysis methods with them. There were 26 patients in the study. Total censoring ratio is 53.85%.

Confidence intervals for survival function in Proportional odds regression model

We first have taken the goodness of fit for the survival times which comes from a log-logistic distribution. To do this, a common and useful technique for checking the validity of a parametric model is to embed it in larger parametric model and use, e.g., the likelihood ratio test to check whether the reduction to the actual model is valid; for applications in survival analysis [8,19].

Secondly, the tests of the survival times come from a log-logistic distribution. We use the Kolmogorov-Smirnov type test. The test statistic result is $D_{26} = 0.1251$. This further indicates that a log-logistic distribution is a reasonable one.

Thirdly, one use of the Nelson-Aalen or Kaplan-Meier estimators for survival data is to check graphically whether the survival time will appear to follow a certain parametric distribution; in fact, this was the rationable behind the estimator in [20] original paper. For the log-logistic distribution with log odds function is $\log\left\{\frac{F(t)}{S(t)}\right\} = \alpha \log \lambda + \alpha \log t$ so that log odds plotted against log t should yield an approximately straight line for the log-logistic distribution This result is given in (Figure 1a) for the ovarian cancer data. The curve is roughly linear, suggesting that a model may be appropriate. In the same Figure, the corresponding log odds estimates (straight lines) based on log-logistic distribution are added the graph and it can seen for approximate the Kaplan-Meier estimates quite good too. In (Figure 1b), QQ plot support same results in (Figure 1a). In (Figure 1c) Baseline Hazard function shows unimodal shape. Also in (Figure 1d), proportional odds regression hazard is a increasing function for risk score function of age.

As a result, (Figure 1d) summarizes that when ages and survival times increases hazard function increases in same direction.

We have also fit the Exponential Hazard Regression (HER), Proportional odds Hazard Regression (POR), Log-Normal Hazard Regression (LNR) and Weibull Hazard Regression (WHR) models to the data in ovarian cancer study. The log likelihood, AIC, AICc and BIC values for each model are given in (Table 1). From the table, we can see that the proportional odds regression model provides the best fits for the data.

As we can see from (Table 1), Proportional odds regression values are smallest for all criterions. Thus we can easily say, Proportional odds regression model is better than the other models.

Therefore we can fit the Proportional odds regression model and the results are given in (Table 2). From the (Table 2), we can see that all three parameters are very significant.

For calculating the confidence intervals of survival functions estimated variance-covariance matrix is given below

$$V\left(\hat{\hat{a}}\right) = \begin{bmatrix} 9.3920 & 0.2210 & -0.1519 \\ 0.2210 & 0.0589 & -0.0031 \\ -0.1519 & -0.0031 & 0.0025 \end{bmatrix}.$$
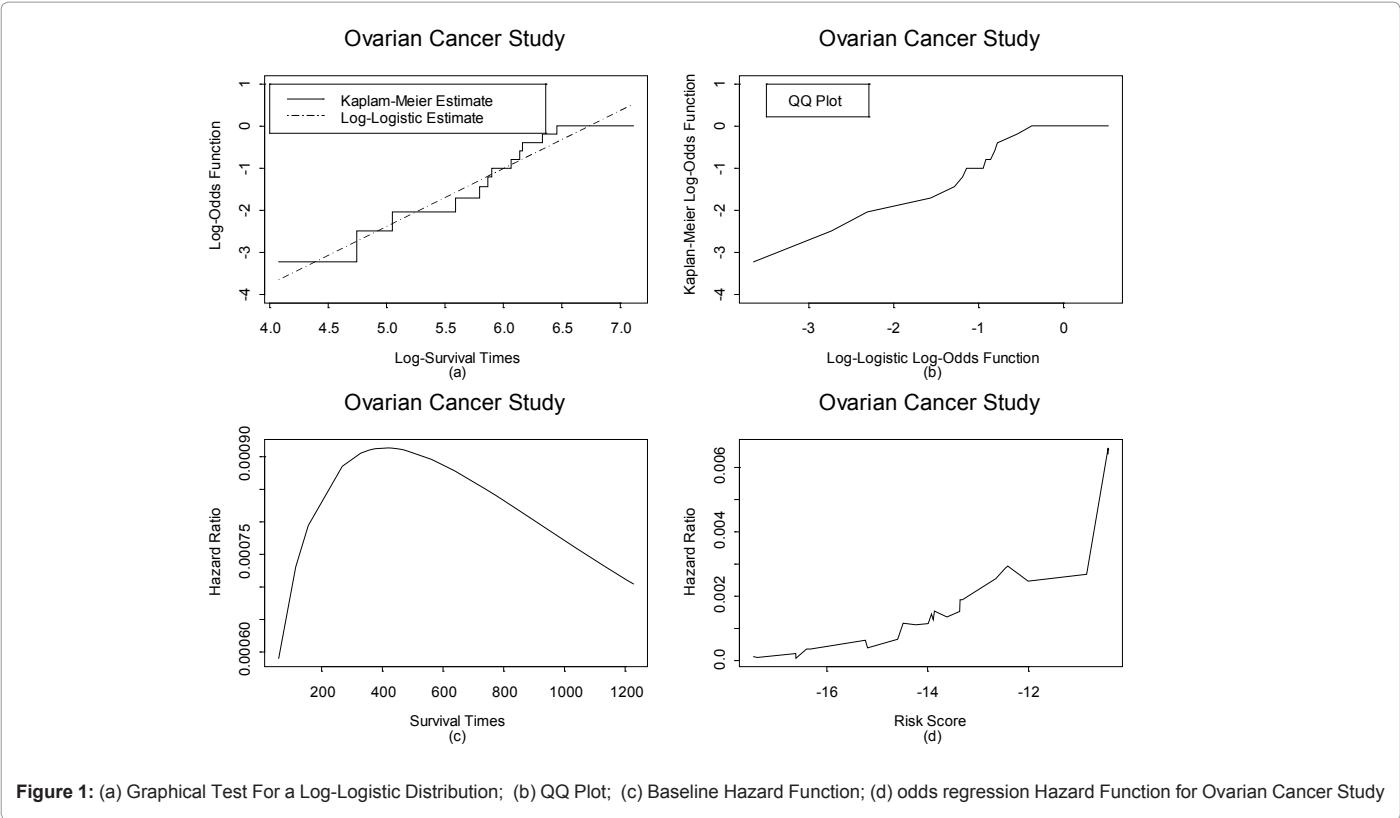
Some calculations of confidence intervals for survival function in the log odds regression model is given in (Table 3). Survival probabilities estimates and 95% confidence intervals are also given (Figure 2).

Approximate 95% confidence limits are obtained using the risk score function and log odds transformation approach.

Figure 2 shows the log odds regression survival function estimate for ovarian cancer with approximate 95% confidence intervals using (18). Because of small sample data the confidence interval results are little bigger than expected. However we can easily say, the result is quite good.

Simulation study

In order to Ovarian Cancer Data has small sample; to see the

Figure 1: (a) Graphical Test For a Log-Logistic Distribution; (b) QQ Plot; (c) Baseline Hazard Function; (d) odds regression Hazard Function for Ovarian Cancer Study

behavior of Proportional odds regression on big samples, a simulation study is studied in this sub-section. Simulation study data has 100 variable value and obtained as follows:
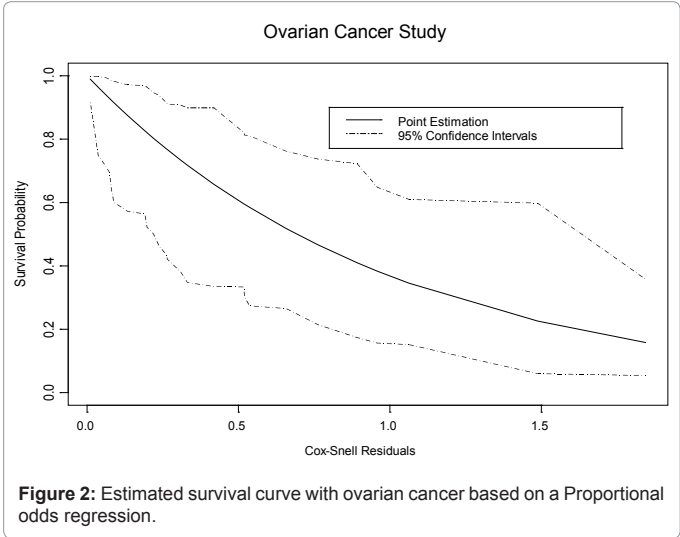
Step 1: Mean, variance and standard deviation for age variable in Ovarian Cancer Data is calculated. ($\hat{\sigma}$ =10.10036, $\hat{\mu}$ = 56.16544)

Step 2: 100 age variable values has been simulated from normal distribution. Here Age = $z\,\hat{\sigma} + \hat{\mu}$ where $z \sim N(0,1)$.

Step 3: 100 random variable value has simulated with uniform distribution in (0,1) interval.

Table 1: Results of Fitting Parametric Models to the Ovarian Cancer Data.

Model	Log Likelihood Value	AIC	AICc	BIC
EHR	-91.7793	185.9986	185.7253	186.8167
POR	-89.5509*	183.1018*	183.6235*	185.6180*
LNR	-89.7349	183.4698	183.9915	185.9860
WHR	-90.0012	184.0024	184.8241	186.5186

Table 2: Results of Proportional odds regression Model to the Ovarian Cancer Data.

Parameter	Value	Std. Err.	z-Test	p-value
$\hat{\beta}_0$ (= $\hat{\alpha} \log \hat{\lambda}$)	-25.9330	3.065	-22.85	0.000
$\hat{\beta}_1$ (= $\log \hat{\alpha}$)	0.8003	0.243	3.30	0.000
$\hat{\beta}_2$ (= $\hat{\alpha}\hat{\theta}$)	0.1975	0.050	3.96	0.001

Figure 2: Estimated survival curve with ovarian cancer based on a Proportional odds regression.

Step 4: With the help of parameter estimation values of Ovarian Cancer Data, survival times is simulated. To do this; $\hat{R}_i = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 *$ Age is evaluated with risk function. Survival time is simulated with the formula $t_i = \left\{ \dfrac{\hat{\alpha}(1 - u_i)e^{-\hat{R}_i}}{u_i} \right\}^{1/\hat{\alpha}}$. Where, $\hat{\alpha}$ =2.22611 and $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 *$ are given as in (Table 2).

Step 5: 100 random variable value is simulated with Bernoulli distribution which has mean value (successful probability)

Figure 3: (a) Graphical Test For a Log-Logistic Distribution; (b) QQ Plot; (c) Baseline Hazard Function; (d) odds regression Hazard Function for Simulation Study.
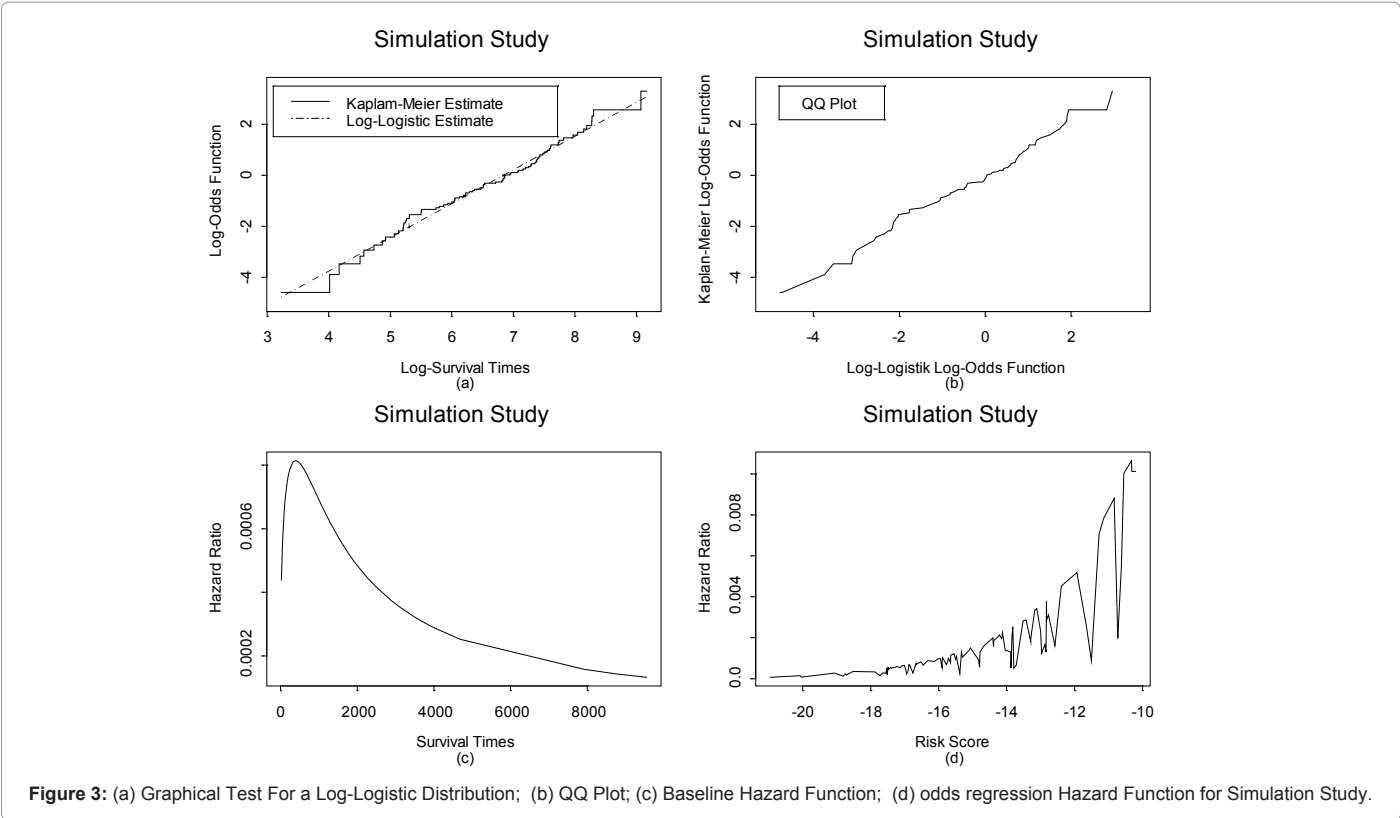
Time	age	$se(\hat{R}_i)$	95% Confidence Intervals			95% Confidence Intervals		
			$\hat{R}_{low(i)}$	R_i	$\hat{R}_{upp(i)}$	$\hat{S}_{low(i)}$	$S(t_i;y_i)$	$\hat{S}_{upp(i)}$
59	72.3315	0.752	-12.32502	-10.85052	-9.376026	0.750	0.929	0.983
115	74.4932	0.831	-12.05233	-10.42368	-8.795035	0.275	0.660	0.908
156	66.4658	0.588	-13.16162	-12.00874	-10.85586	0.602	0.827	0.938
421	53.3644	0.691	-15.94933	-14.59569	-13.24205	0.644	0.875	0.964
431	50.3397	0.793	-16.74785	-15.19293	-13.63802	0.718	0.923	0.983
448	56.4301	0.608	-15.18141	-13.99035	-12.79929	0.502	0.769	0.916
464	56.9370	0.597	-15.05972	-13.89026	-12.72080	0.463	0.735	0.900
475	59.8548	0.552	-14.39591	-13.31412	-12.23233	0.335	0.597	0.814
477	64.1753	0.554	-13.54678	-12.46101	-11.37525	0.174	0.385	0.650
563	55.1781	0.638	-15.48886	-14.23756	-12.98627	0.423	0.719	0.899
638	56.7562	0.600	-15.10292	-13.92596	-12.74899	0.304	0.586	0.821
744	50.1096	0.802	-16.80984	-15.23837	-13.66690	0.437	0.789	0.947
769	59.6301	0.554	-14.44461	-13.35849	-12.27237	0.152	0.347	0.611
770	57.0521	0.594	-15.03232	-13.86753	-12.70274	0.215	0.468	0.738
803	39.2712	1.256	-19.84033	-17.37847	-14.91662	0.696	0.964	0.997
855	43.1233	1.085	-18.74517	-16.61785	-14.49054	0.565	0.916	0.989
1040	38.8932	1.273	-19.94846	-17.45311	-14.95777	0.573	0.942	0.995
1106	44.6000	1.022	-18.32942	-16.32627	-14.32312	0.383	0.821	0.971
1129	53.9068	0.674	-15.80995	-14.48859	-13.16723	0.157	0.411	0.724
1206	44.2055	1.039	-18.44022	-16.40417	-14.36811	0.348	0.804	0.969
1227	59.5890	0.555	-14.45356	-13.36660	-12.27965	0.060	0.159	0.359
268	74.5041	0.831	-12.05099	-10.42153	-8.792072	0.054	0.227	0.600
329	43.1370	1.085	-18.74130	-16.61515	-14.48900	0.916	0.989	0.999
353	63.2192	0.546	-13.72033	-12.64980	-11.57927	0.336	0.596	0.812
365	64.4247	0.557	-13.50280	-12.41177	-11.32073	0.266	0.520	0.763
377	58.3096	0.571	-14.73922	-13.61923	-12.49924	0.523	0.771	0.912

Table 3: 95% Confidence Intervals for Survival Probabilities of Proportional odds regression Model to the Ovarian Cancer Data.

p=0.80. Therefore 20% censored survival times is identified using these variables.

The result for simulation data in graphs is given in (Figure 3). In (Figure 3a), the curve is quite linear. In the same Figure, the corresponding log odds estimates (straight lines) based on log-logistic distribution are added the graph and it can seen for approximate the Kaplan-Meier estimates are very good. In (Figure 3b), QQ plot support same results in (Figure 3a). In (Figure 3) Baseline Hazard function shows unimodal shape. Also in (Figure 3d), proportional odds regression hazard is a increasing for risk score function of age.

Table 4 results shows that Proportional odds regression model has the smallest criterion values than the other models. We can easily say that it is the best model above others. In Table 5, estimation of parameters is given and we can say that all estimations are very significant.

For calculating the confidence intervals of survival functions estimated variance-covariance matrix is given below:

$$V\left(\hat{\hat{a}}\right) = \begin{bmatrix} 1.116442 & 0.010860 & -0.01848 \\ 0.010860 & 0.00874 & -0.00016 \\ -0.01848 & -0.00016 & 0.00032 \end{bmatrix}.$$

Model	Log Likelihood Value	AIC	BIC
EHR	-644.8942	1291.788	1294.3936
POR	-623.6054*	1251.211*	1256.4211*
LNR	-626.3671	1256.734	1261.9445
WHR	-643.6997	1291.399	1296.6097

Table 4: Results of Fitting parametric models to the simulated data.

$\hat{\beta}_0 \; (=\hat{\alpha}\log\hat{\lambda})$	-28.9309	1.057	-27.381	0.000
$\hat{\beta}_1 \; (=\log\hat{\alpha})$	0.8687	0.093	9.294	0.000
$\hat{\beta}_2 \; (=\hat{\alpha}\hat{\theta})$	0.2218	0.018	12.492	0.000
Parameter	Value	Std. Err.	z-Test	p-value

Table 5: Results of Proportional odds regression Model to the Simulated Data.

Figure 4: Estimated survival curve with simulated data based on a proportional odds regression.
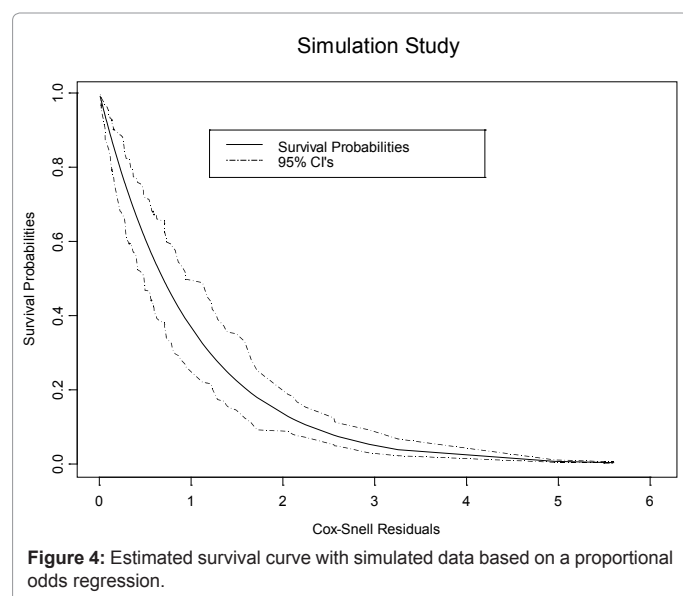
In (Figure 4), intervals of simulation data are quite small. Cox-Snell residuals intervals gave robust intervals compared to real time data. Thus we can easily say that when data sample gets bigger, expected intervals will be smaller on survival structural data sets.

Discussion

Many statistical investigations can occur both estimation and hypothesis testing. Estimation can be made in two different types. One is point estimation and the other is interval estimation. Both point and interval estimations can be achieve with an estimator. Interval estimation is generally called confidence interval estimation and naturally the estimators are also called confidence interval estimators.

Survival function may be the most important function in survival analysis or reliability analysis. Probability of living longer than t time is an important issue for both doctors and patients or patient relatives. Researching for factors which affects this issue is also important for determining risk factor function variables on survival times. It is necessary to search influence of hazard function for survival analysis. Both point estimation and confidence interval estimation of the survival function may be achieved by fitting parametric distributions. Semi-parametric proportional hazard model is known Cox regression model. In the Cox regression model confidence interval estimation of survival function has studied by Link [6,7]. Her idea more recently extended by Alakus et al. [1] to the exponential distribution and Alakus et al. [2] to exponential proportional hazard model, respectively. Weibull proportional hazard function is also investigated by Alakuş [3].

For this reason, in this study we offer a new confidence interval with transformed log odds in symmetric type in proportional odds regression model. Proposed approach is studied with real time data and simulation data. Results were quite good enough. When sample size getting larger, confidence intervals were getting tighter. Model selection criterions Log-Likelihood, AIC and BIC gave the best results for proposed model. Confidence intervals are narrower than other studied models. At the thought real time data has small sample, a larger data set with simulation data also studied. Results are more robust than small sample. Based on the theorem of large numbers law; Ŝ(t;y) will be converges in probability to S(t;y). This summarizes us when sample size is quite enough, distribution of risk function approximate to normal distribution and this becomes Ŝ(t;y) closer to S(t;y). This is what we are investigating for.

The investigated confidence intervals may be extended to log-normal hazard model. This problem will be investigated by the forthcoming studies.

References

1. Alakuş K, Öner Y, Tunç T (2007a) Tamamlanmış ve Sansürlü Örneklemlerde Üstel Dağılımın Sağ Kalım Fonksiyonu için Güven Aralığı Metotlarının Karşılaştırılması (in Turkish) Beşinci İstatistik Kongresi, Antalya 449-459.

2. Alakuş K, Tunç T, Öner Y (2007b) Üstel Orantılı Hazard Regresyon Modelinde Sağ Kalım Fonksiyonu için Güven Aralığı Tahmini. İstatistik Araştırma Sempozyumu

3. Alakus K (2010) Confidence Intervals Estimation for Survival Function in Weibull Proportional Hazards Regression Based on Censored Survival Time Data. Science Research Essays 5: 1589-1594.

4. Kaplan EL, Meier P (1958) Nonparametric Estimation from Incomplete Observations J Amer Statist Ass 53: 457-481.

5. Thomas DR, Grunkemeier GL (1975) Confidence Interval Estimation of Survival Probabilities for Censored Data J Amer Statist. Ass 70: 865-871.

6. Link CL (1984) Confidence Intervals for The Survival Function Using Cox's Proportional Hazard Model With Covariates Biometrics 40: 601-610.

7. Link CL (1986) Confidence Intervals for the Survival Function in the Presence of Covariates. Biometrics 42: 219-220.

8. Cox, D.R., Oakes, D.(1984). Analysis of Survival Time Data, Chapman and Hall, London.

9. Collett D (1994). Modelling Survival Data in Medical Research. Chapman and Hall, London, (1994).

10. Alakus K (1997) Contributions to Statistical Estimation and Testing in Survival Analysis Problems. Unpublished PhD Thesis, University of Strathclyde, UK.

11. Marubini E, Valsecchi MG (1995) Analsing Survival Data from Clinical Trials and Observational Studies. New York: John Wiley & Sons.

12. Crowder MJ, Kimber AC, Smith RL, Sweeting TJ (1991) Statistical Analysis of Reliability Data, Chapman and Hall, London.

13. Bennett S (1983a) Analysis Survival Data by the Proportional odds Model. Statistics in Medicine 2: 273-277.

14. Bennett, S. (1983b). Log-Logistic Regression Models for Survival Data. Applied Statistics, 32, 165-71.

15. Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19: 716–723.

16. Schwarz GE (1978) Estimating the Dimension of a Model. Annals of Statistics 6: 461–464.

17. Edmunson JH, Fleming TR, Decker DG, Malkasian GD, Jefferies JA et al. (1979) Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma vs. minimal residual disease. Cancer Treatment Reports 63: 241-247.

18. Statistical Sciences (1995) S-PLUS Version 3.3 Supplement. StatSci: Seattle.

19. Kalbfleisch JD, Prentice RL (1991) The Statistical Analysis of Failure time data. (2nd edn), Wiley, New York.

20. Nelson W (1969) "Hazard Plotting For Incomplete Failure Data".J Qual Technol 1: 27-52.