Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 49 (1) (2020), 458 – 477
DOI : 10.15672/hujms.617303

Research Article

Proportional hazards model under ranked set sampling scheme using censored data of coronary heart disease

Nursel Koyuncu*, Nihal Ata Tutkun

Department of Statistics, Hacettepe University, Beytepe, Turkey

Abstract

The proportional hazards model is one of the most common model for analyzing survival data. Only proportional hazards assumption is required to apply this model. Using appropriate sampling methods is an important part of modelling data and estimation of parameters. In literature there is a few studies based on sampling methods in survival analysis and most of them are related with non-parametric estimations of survival functions, sample size calculation etc. The main innovation of our approach is to examine the sampling methods for the proportional hazards model. This paper describes usage of ranked set sampling design in the proportional hazards model. In order to analyze the performance of our methods, we use a real data and conduct a simulation study. We conclued that ranked set sampling is more efficient than simple random sampling.

Mathematics Subject Classification (2010). 62N01, 62D05

Keywords. Cox regression, likelihood function, censored data, ranked set sampling

1. Introduction

Survival analysis has several applications especially in medical sciences. The main outcome is survival time which is defined as the time to the occurrence of an interested event. This event can be the death, disease, relapse or recovery. In survival analysis some observations have not experienced the event and their exact survival times are unknown. These observations are called as censored observations and this feature makes survival analysis necessary and different than classical statistical methods. The relation between the survival time and covariates are investigated using survival models. The broadly applicable and the most widely used survival model is proportional hazards model (PHM) which was suggested by Cox [11].

Follow up time, end points, interested event, covariates and sample size are the key points of PHM as well as survival analysis. The importance that can be attached to the results from a survival analysis depends on the selection and number of observations included [3]. An important question is how large the sample size should be to come to accurate conclusions with respect to the effect of covariates [24].

*Corresponding Author.
Email addresses: nkoyuncu@hacettepe.edu.tr (N. Koyuncu), nihalata@hacettepe.edu.tr (N. Ata Tutkun)
Received: 09.09.2019; Accepted: 24.01.2020

The sample size and power calculation studies in PHM are very rare [9,15,27,29]. This may occur because of the structure of survival data and limited access to a data with censored observations. However, in recent years the capacity of the medical databases increases and this makes the applications difficult in statistical packages. For this reason, determining the sample size and sampling methods become important.

Sampling methods which we use in statistical analysis are very important and affect our estimates. Ranked Set Sampling (RSS), firstly proposed by McIntyre [23], takes into account the field investigator's professional knowledge and judgment to select sample. The use of professional judgment in the process of selecting sampling is a powerful incentive to use RSS. This results in better estimates of the mean as well as improved performance of many statistical procedures instead of using simple random sampling (SRS). Many researchers introduced new sampling procedures through modifying original RSS. We can list some references as [1, 2, 17, 18, 33]. In recent years, Ozturk [25] introduced a new sampling design which is similar to RSS design with a clear difference that rankers are allowed to declare any two or more units are tied in ranks whenever the units can not be ranked with high confidence. Frey [13] considered a modification of RSS that allows the ranker to declare ties and proposed several different nonparametric mean estimators that incorporate the tie information. Mahdizadeh and Zamanzade [19] developed a kernel-based estimator of a dynamic reliability measure for use with independent ranked set samples. Mahdizadeh and Zamanzade [20] dealed with constructing a confidence interval for the reliability parameter using ranked set sampling. Mahdizadeh and Zamanzade [21] introduced a new design, called multistage pair RSS. Gemayel et al. [14] compared the statistical precision of SRS with balanced RSS in an inventory valuation scenario. Zamanzade and Mahdizadeh [34] studied the properties of the maximum likelihood estimator of the population proportion in RSS with extreme ranks.

It is also well known that using RSS design is better than SRS design when the ranking of observation is easy. Survival models are based on discrete-time or continuous-time. In the literature using RSS design in discrete-time survival models were studied firstly by Ata Tutkun et.al.[4]. For contionuous-time survival models, Samawi [26] proposed new RSS schemes to associate with survival time in accelerated failure time models. In this paper, our aim is to use RSS design in PHM which is the most popular continuous-time survival model. Moving this direction our aim is studying PHM under different sampling designs and compare the efficiency.

The rest of the paper is organized as follows: Section 2 gives a brief review of PHM. The sampling designs are introduced in Section 3. We propose PHM under RSS in the next section. A simulation study is conducted using a real data coronary heart disease in Section 5. And finally, we summarize our results in Section 6.

2. Proportional hazards model

The PHM [11] is the most common model used for analyzing survival time data. It is a model which describes the relation between the event incidence, as expressed by the hazard function and a set of covariates [6]. The model has a proportional hazards assumption and survival times are not assumed to follow a particular statistical distribution. The idea of the PHM is to define hazard level as a dependent variable which is being explained by the time-related component (so called baseline hazard) and covariate-related component [7]. Hazard function for the PHM is given by

$$h(t) = h_0(t)exp(\boldsymbol{\beta}'\boldsymbol{x}) \tag{2.1}$$

where $\mathbf{x}$ is the matrix including categorical variables or continuous measurements of each unit, $h_0(t)$ is the baseline hazard function, and $\boldsymbol{\beta}$ is a $[p * 1]$ vector of unknown parameters.

The ordered failure times are denoted by $t_1 < \cdots < t_r$ and the set of units who are at risk at time $t_{(j)}$ are denoted by $R(t_{(j)})$. Then, the likelihood for PHM is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}{\sum\limits_{\ell \epsilon R(t_j)} exp(\boldsymbol{\beta}'\boldsymbol{x}_\ell)} \tag{2.2}$$

where $\mathbf{x}_j$ is the vector of covariates for the unit who fails at the j^{th} ordered failure time, $t_{(j)}$.

In censored data set, there is a possibility of more than one failed observation or censored observation at a given time and such observations are called as tied observations (observations with the same survival time). The Breslow approximation [8], Efron approximation [12] and Kalbfleisch and Prentice [18] exact expression are proposed to handle the tied observations in PHM. Breslow [8] proposed the following approximation to the likelihood function,

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta}'\boldsymbol{s}_j)}{[\sum\limits_{\ell \epsilon R(t_j)} exp(\boldsymbol{\beta}'\boldsymbol{x}_\ell)]^{d_j}} \tag{2.3}$$

where d_j is the event indicator taking 1 if the event has occurred and 0 otherwise. In Eq. 2.3, the $\boldsymbol{s}_j$ is the vector of sums of each of the p covariates for those units who die at the j^{th} failure time, $t_{(j)}$, $j = 1, \ldots, r$. If there are d_j failures at $t_{(j)}$, the h^{th} element of $\boldsymbol{s}_j$ is $s_{hj} = \sum_{k=1}^{d_j} x_{hjr}$, where x_{hjr} is the value of the h^{th} variable, $h = 1, \ldots, p$, for the k^{th} of d_j units, $k = 1, \ldots, d_j$ who fail at the j^{th} failure time [10].

Efron [12] proposed

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta}'\boldsymbol{s}_j)}{\prod_{k=1}^{d_j}[\sum\limits_{\ell \epsilon R(t_{(j)})} exp(\boldsymbol{\beta}'\boldsymbol{s}_\ell) - (k-1)d_k^{-1} \sum\limits_{\ell \epsilon D(t_{(j)})} exp(\boldsymbol{\beta}'\boldsymbol{x}_\ell)]} \tag{2.4}$$

as an approximate likelihood for PHM, where $D(t_{(j)})$ is the set of all units who fails at time $t_{(j)}$. The exact likelihood is given by,

The exact likelihood is given by,

$$L(\beta) = \prod_{j=1}^{r} [\int_0^\infty \prod_{i=1}^{d_j} (1 - exp(-\frac{exp(\boldsymbol{\beta}'\boldsymbol{s}_j)}{\sum\limits_{\ell \epsilon R(t_{(j)})} exp(\boldsymbol{\beta}'\boldsymbol{x}_\ell)} * t)) * exp(-t)dt]. \tag{2.5}$$

The Breslows likelihood is an adequate approximation when the number of tied observations at any time is not too large. Efrons approximation is a closer approximation to the appropriate likelihood function than due to Breslow, although in practice, both approximations often give similar results [10]. However, if time is not limited, one should consider choosing an exact method that can provide better fit statistics and more efficient parameter estimates [7]. When there no ties, all the approximations reduce to the likelihood function in Eq. 2.2 or Eq. 2.3.

3. Sampling designs

In this section, to evaluate the performance of PHM in different designs we consider SRS and RSS. Also the notations of these designs are given. In RSS designs there are two approaches toward concominant based RSS depending on whether the actual measurements of the concomitant variable is available or not. Using actual quantifications of the concomitant variable onto estimation process may improve upon the parameter estimation once more. We can list some important studies in literature as: Ashour and Abdallah [5] showed that using the concomitant information can lead to improved estimation of the cumulative distribution function under RSS set up. Frey [13] provided that when rankings are done using a covariate, the standard RSS mean estimators no longer make efficient use of the available information. Zamanzade and Vock [30] proposed a nonparametric variance estimator when RSS are applied by measuring a concomitant variable. Zamanzade and Mahdizadeh [32] proposed a new estimator of distribution function when RSS is done by using a concomitant variable. Zamanzade and Mahdizadeh [31] proposed a new estimator for the population proportion using a concomitant-based RSS scheme. In this study, we assume that we have a bivariate data contains study and concomitant variables and the judgement ranking is done using a concomitant variable.

3.1. SRS design

In sampling literature generally Y is the study variable however in survival analysis the interested variable is survival time and represented by T. So, let T be the study; X be auxiliary variable associated with each unit of the population. A sample of size n is drawn without replacement from the population. Let $(X_1, T_1), (X_2, T_2), \dots, (X_n, T_n)$ denote the observed values of X and T. Moreover let μ_t, μ_x, $\bar{t}_{SRS} = \frac{\sum_{i=1}^{n} T_i}{n}$, $\bar{x}_{SRS} = \frac{\sum_{i=1}^{n} X_i}{n}$ be the population and sample means of study and auxiliary variable respectively.

3.2. RSS design

RSS is introduced by McIntyre [23]. RSS design can be described as follows:

(1) Select a simple random sample of size n^2 units from the target population and divide them into n samples each of size n.
(2) Rank the units within each sample in increasing magnitude using personal judgment, eye inspection or based on a concomitant variable.
(3) Select the i^{th} ranked unit from the i^{th} sample.
(4) Repeat steps (1) through (3) m times if needed to obtain a ranked set sample of size $N = nm$.

Let $(X_1, T_1), (X_2, T_2), \dots, (X_n, T_n)$ be a simple random sample of size n, then the measured RSS units are denoted by $(X_{(i)j}, T_{[i]j}), i = 1, \dots, n, j = 1, \dots, m$ where $(X_{(i)j}, T_{[i]j})$ is the i^{th} ranked unit from the j^{th} cycle of auxiliary variable and study variable respectively, () and [] indicate that the ranking of X is perfect and ranking of T has errors. The sample means of variables can be defined as in RSS:

$$\bar{t}_{(RSS)} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{i=1}^{n} T_{[i]j}, \tag{3.1}$$

$$\bar{x}_{(RSS)} = \frac{1}{nm} \sum_{j=1}^{m} \sum_{i=1}^{n} X_{(i)j}. \tag{3.2}$$

4. Suggested methods for PHM under different sampling designs

In this section we try to combine PHM models with sampling designs described in above sections. Using Breslow (in Eq. 2.3), Efron (in Eq. 2.4), and Exact (in Eq. 2.5), methods in different sampling designs we get following models:

$$L(\boldsymbol{\beta}_{(i)}) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta}'_{(i)} \boldsymbol{s}_{j(i)})}{[\sum_{\ell \epsilon R(t_j)} exp(\boldsymbol{\beta}'_{(i)} \boldsymbol{x}_{\ell(i)})]^{d_j}} \tag{4.1}$$

$$L(\boldsymbol{\beta}_{(i)}) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta}_{(i)}' \boldsymbol{s}_{j(i)})}{\prod_{k=1}^{d_j}[\sum\limits_{\ell \epsilon R(t_{(j)})} exp(\boldsymbol{\beta}_{(i)}' \boldsymbol{s}_{\ell(i)}) - (k-1)d_k^{-1} \sum\limits_{\ell \epsilon D(t_{(j)})} exp(\boldsymbol{\beta}_{(i)}' \boldsymbol{x}_{\ell(i)})]} \tag{4.2}$$

$$L(\boldsymbol{\beta}_{(i)}) = \prod_{j=1}^{r} [\int_0^\infty \prod_{i=1}^{d_j} (1 - exp(-\frac{exp(\boldsymbol{\beta}_{(i)}' \boldsymbol{s}_{j(i)})}{\sum\limits_{\ell \epsilon R(t_{(j)})} exp(\boldsymbol{\beta}_{(i)}' \boldsymbol{x}_{\ell(i)})} * t)) * exp(-t)dt]. \tag{4.3}$$

where $(i) = SRS, RSS$ representing different sampling designs.

5. Simulation study

The simulation study is performed to assess the properties of PHM under different sampling schemes. We consider a real data set of Whitehall I ([22], [28]) which is a prospective, cross-sectional cohort study of male British Civil Servants employed in London. We focus on a cohort of 17 260 male Civil Servants aged 4064 years with complete 10-year follow-up. Time to death from coronary heart disease (CHD) is treated as a censored survival-time outcome. In this study, age and systolic blood pressure (sysbp) are used as covariates.

The samples were generated from Whitehall I data. To evaluate the sample size effect, ordering variable effect and censoring effect, the sample sizes $n = 20, 50, 100, 200$ and censoring rates 10% , 30%, 60% were chosen.

We consider Breslow, Exact and Efron approximations of likelihood functions used in PHM to estimate beta parameters. MSE value of each parameter estimation is computed using the equations given below:

$$MSE = \frac{\sum_{k=1}^{10000}(\hat{\beta}_{age(i)k} - \beta_{age})}{10000}, MSE = \frac{\sum_{k=1}^{10000}(\hat{\beta}_{sysbp(i)k} - \beta_{sysbp})}{10000} \tag{5.1}$$

where $\hat{\beta}_{\ell(i)}$ represents $(i) = SRS, RSS$ and $\ell = age, sysbp$. The percentage of relative efficiency (PRE) is calculated as,

$$PRE = \frac{MSE(\hat{\beta}_{\ell(SRS)})}{MSE(\hat{\beta}_{\ell(RSS)})} * 100 \tag{5.2}$$

The computed values are given in tables.

6. Simulation results

In simulation study, to select the samples we consider age and sysbp as ordering variables respectively. In Table 1-6 we summarized the results according to age variable and in Table 7-12 we summarized the results according to sysbp. In Table 1-3-5-7-9-11, parameter estimations of two covariates were estimated by using MSE and PRE values under different sampling methods. In Table 2-4-6-8-10-12 we compared the models according to MSE and PRE values. The results of Breslow method are given in Table 1-2-7-8, Efron Method in Table 3-4-9-10 and Exact Method in Table 5-6-11-12. The following results were concluded from the tables:

For ordering variable age:
We can see that the parameter estimations of age have the minimum MSE values for different sample sizes, censoring rates and different likelihood approximation methods. The PRE values of Breslow, Efron and Exact methods under small sample size $n = 20$, 60% censoring rate are 7406.202; 7405.892 and 8020.127 respectively (see Table 1-3-5). From these values we can say that parameters estimations of RSS method are 74-80 times more efficient than SRS method.

For the parameter estimations of sysbp the PRE values of Breslow, Efron and Exact methods under small sample size $n = 20$, 60% censoring rate are 325.5347; 325.3925 and 345.2115 respectively (see Table 1-3-5). From these values we can conclude that parameter estimations of sysbp under RSS are approximately 3 times more efficient than SRS. However for the other cases the parameter estimations of RSS are similar with SRS. When the look at the model efficiency in Tables 2-4-6, we can see that RSS method is better than SRS under all cases. Especially in small sample size $n = 20$, 60% censoring rate, RSS is approximately 60 times more efficient than SRS. In all other sample sizes, under 60% censoring rate, RSS is approximately 2 times more efficient than SRS.

As a result, RSS method is better than SRS for the efficiency of parameter estimations and model. Especially for small sample size and high censoring rate we advise using RSS method instead of SRS for Cox regression model.

For ordering variable sysbp:

We can see that the parameter estimations of sysbp have the minimum MSE values for different sample sizes, censoring rates and different likelihood approximation methods. The PRE values of Breslow, Efron and Exact methods under small sample size $n = 20$, 60% censoring rate are 694,7909; 694,6371 and 736,9444 respectively (see Table 7-9-11). From these values we can say that parameters estimations of RSS method are approximately 7 times more efficient than SRS method. When the ordering variable is age we can see approximately 3 times efficiency. So we can conclude that which variable we consider as ordering covariate, it affects the parameter of estimation mostly. For the parameter estimation of age the PRE values of Breslow, Efron and Exact methods under small sample size $n = 20$, 60% censoring rate are 3106.421; 3104.185 and 3361.916 respectively (see Table 7-9-11). From these values we can conclude that parameter estimations of age under RSS are approximately thirty times more efficient than SRS. However for the other cases the parameter estimations of RSS are similar with SRS. When the look at the model efficiency in Tables 8-10-12, we can see that RSS method is better than SRS under all cases. Especially in small sample size $n = 20$, 60% censoring rate, RSS is approximately 30 times more efficient than SRS.

7. Conclusion

In this paper, we have studied proportional hazards model under simple random sampling and ranked set sampling. In the literature there are a few studies combining sampling schemes and survival analysis. And these studies are mostly related with sample size calculation. Apart from these studies this paper has proposed a more efficient sampling method for PHM. Moving this direction, we have used SRS and RSS for PHM and conducted a simulation study.

In real data example, we have concluded that RSS method is better than SRS for parameter estimations and the efficiency of the model. Especially for small sample size and high censoring rate we have advised to use RSS method instead of SRS for PHM. Morover we can see approximately three times efficiency when the ordering variable is sysbp whereas there is approximately seven times efficiency when the ordering variable is age.

Acknowledgment. The authors are thankfull to learned referees for their valuable comments to improve our paper.

Table 1. MSE and PRE values parameter estimation under Breslow Method (ordering variable: age).

		MSE and PRE values of Beta of Age		**MSE and PRE values of Beta of Sysbp**	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,00511	**0,004623***	0,000263	0,000269
	PRE	100	**110,5309***	100	97,8635
30%	**MSE**	0,004207	**0,002704***	0,000309	**0,000300***
	PRE	100	**155,5567***	100	**102,9521***
60%	**MSE**	0,229129	**0,003094***	0,002092	**0,000643***
	PRE	100	**7406,202***	100	**325,5347***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	**0,003054***	0,000102	**0,000101***
	PRE	100	**103,9296***	100	**100,6777***
30%	**MSE**	0,001837	**0,001365***	0,000095	0,000098
	PRE	100	**134,656***	100	97,79357
60%	**MSE**	0,001856	**0,000692***	0,000153	**0,000152***
	PRE	100	**268,255***	100	**100,6675***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002786	**0,002715***	0,000071	**0,000069***
	PRE	100	**102,6356***	100	**101,9596***
30%	**MSE**	0,001365	**0,001109***	0,000053	0,000054
	PRE	100	**123,0873***	100	97,87396
60%	**MSE**	0,000902	**0,000348***	0,000066	0,000067
	PRE	100	**259,1784***	100	97,74972
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002685	**0,002610***	0,000057	0,000057
	PRE	100	**102,8569***	100	98,76765
30%	**MSE**	0,001152	**0,001032***	0,000037	**0,000036***
	PRE	100	**111,5761***	100	**100,9117***
60%	**MSE**	0,000511	**0,000240***	0,000032	0,000033
	PRE	100	**213,1305***	100	99,29733

Table 2. MSE and PRE values of model under Breslow Method (ordering variable: age)

Sample Size	Censoring Rate		Sampling Method	
			SRS	RSS
n=20	**10%**	**MSE**	0,002686	**0,002446***
		PRE	100	**109,84***
	30%	**MSE**	0,002258	**0,001502***
		PRE	100	**150,30***
	60%	**MSE**	0,115611	**0,001868***
		PRE	100	**6188,39***
n=50	**10%**	**MSE**	0,001638	**0,001578***
		PRE	100	**103,83***
	30%	**MSE**	0,000966	**0,000731***
		PRE	100	**132,20***
	60%	**MSE**	0,001005	**0,000422***
		PRE	100	**238,01***
n=100	**10%**	**MSE**	0,001429	**0,001392***
		PRE	100	**102,62***
	30%	**MSE**	0,000709	**0,000582***
		PRE	100	**121,92***
	60%	**MSE**	0,000484	**0,000208***
		PRE	100	**233,06***
n=200	**10%**	**MSE**	0,001371	**0,001334***
		PRE	100	**102,77***
	30%	**MSE**	0,000594	**0,000534***
		PRE	100	**111,21***
	60%	**MSE**	0,000272	**0,000136***
		PRE	100	**199,54***

Table 3. MSE and PRE values parameter estimation under Efron Method (ordering variable: age)

		MSE and PRE values of Beta of Age		**MSE and PRE values of Beta of Sysbp**	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,00511	**0,004623***	0,000263	0,000269
	PRE	100	**110,5475***	100	97,81065
30%	**MSE**	0,004207	**0,002704***	0,000309	**0,000301***
	PRE	100	**155,5615***	100	**102,8762***
60%	**MSE**	0,229131	**0,003094***	0,002092	**0,000643***
	PRE	100	**7405,892***	100	**325,3925***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	**0,003052***	0,000102	**0,000101***
	PRE	100	**103,9918***	100	**100,6725***
30%	**MSE**	0,001837	**0,001364***	0,000095	0,000098
	PRE	100	**134,7507***	100	97,7325
60%	**MSE**	0,001857	**0,000692***	0,000153	**0,000152***
	PRE	100	**268,3241***	100	**100,6293***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002786	**0,002712***	0,000071	**0,000069***
	PRE	100	**102,7126***	100	**101,9938***
30%	**MSE**	0,001365	**0,001108***	0,000053	0,000054
	PRE	100	**123,2028***	100	97,8654
60%	**MSE**	0,000902	**0,000348***	0,000066	0,000067
	PRE	100	**259,3025***	100	97,71519
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002684	**0,002607***	0,000057	0,000057
	PRE	100	**102,9624***	100	98,83246
30%	**MSE**	0,001152	**0,001031***	0,000037	**0,000036***
	PRE	100	**111,7165***	100	**100,9352***
60%	**MSE**	0,000511	**0,000240***	0,000032	0,000033
	PRE	100	**213,3664***	100	99,26107

Table 4. MSE and PRE values of model under Efron Method (ordering variable: age)

Sample Size	**Censoring Rate**		**Sampling Method**	
			SRS	**RSS**
n=20	**%10**	**MSE**	0,002687	**0,002446***
		PRE	100	**109,85***
	%30	**MSE**	0,002258	**0,001502***
		PRE	100	**150,29***
	%60	**MSE**	0,115611	**0,001868***
		PRE	100	**6187,71***
n=50	**%10**	**MSE**	0,001638	**0,001577***
		PRE	100	**103,89***
	%30	**MSE**	0,000966	**0,000731***
		PRE	100	**132,28***
	%60	**MSE**	0,001005	**0,000422***
		PRE	100	**238,04***
n=100	**%10**	**MSE**	0,001428	**0,001391***
		PRE	100	**102,69***
	%30	**MSE**	0,000709	**0,000581***
		PRE	100	**122,03***
	%60	**MSE**	0,000484	**0,000208***
		PRE	100	**233,14***
n=200	**%10**	**MSE**	0,001370	**0,001332***
		PRE	100	**102,87***
	%30	**MSE**	0,000594	**0,000534***
		PRE	100	**111,35***
	%60	**MSE**	0,000272	**0,000136***
		PRE	100	**199,73***

Table 5. MSE and PRE values parameter estimation under Exact Method (ordering variable: age)

		MSE and PRE values of Beta of Age		**MSE and PRE values of Beta of Sysbp**	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,005111	**0,004624***	0,000263	0,000269
	PRE	100	**110,5394***	100	97,75003
30%	**MSE**	0,004207	**0,002705***	0,000309	**0,000301***
	PRE	100	**155,5452***	100	**102,8475***
60%	**MSE**	0,248202	**0,003095***	0,00222	**0,000643***
	PRE	100	**8020,127***	100	**345,2115***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	**0,003051**	0,000102	**0,000101***
	PRE	100	**104,0201**	100	**100,6643***
30%	**MSE**	0,001837	**0,001363**	0,000095	0,000098
	PRE	100	**134,8126**	100	97,72936
60%	**MSE**	0,001857	**0,000692**	0,000153	**0,000153***
	PRE	100	**268,2845**	100	**100,6089***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002786	**0,002710***	0,000071	**0,000069***
	PRE	100	**102,7756***	100	**102,0286***
30%	**MSE**	0,001365	**0,001107***	0,000053	0,000054
	PRE	100	**123,3167***	100	97,8892
60%	**MSE**	0,000902	**0,000348***	0,000066	0,000067
	PRE	100	**259,4771***	100	97,69674
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002684	**0,002605***	0,000057	0,000057
	PRE	100	**103,0316***	100	98,91842
30%	**MSE**	0,001151	**0,001030***	0,000037	**0,000036***
	PRE	100	**111,832***	100	**101,0276***
60%	**MSE**	0,000511	**0,000239***	0,000032	0,000033
	PRE	100	**213,6435***	100	99,2739

Table 6. MSE and PRE values of model under Exact Method (ordering variable: age)

Sample Size	Censoring Rate		Sampling Method	
			SRS	RSS
n=20	**%10**	**MSE**	0,002687	**0,002447***
		PRE	100	**109,84***
	%30	**MSE**	0,002258	**0,001503***
		PRE	100	**150,27***
	%60	**MSE**	0,125211	**0,001869***
		PRE	100	**6699,60***
n=50	**%10**	**MSE**	0,001638	**0,001576***
		PRE	100	**103,91***
	%30	**MSE**	0,000966	**0,000730***
		PRE	100	**132,33***
	%60	**MSE**	0,001005	**0,000422***
		PRE	100	**238,00***
n=100	**%10**	**MSE**	0,001428	**0,001390***
		PRE	100	**102,76***
	%30	**MSE**	0,000709	**0,000580***
		PRE	100	**122,14***
	%60	**MSE**	0,000484	**0,000208***
		PRE	100	**233,26***
n=200	**%10**	**MSE**	0,001370	**0,001331***
		PRE	100	**102,94***
	%30	**MSE**	0,000594	**0,000533***
		PRE	100	**111,46***
	%60	**MSE**	0,000272	**0,000136***
		PRE	100	**199,96***

Table 7. MSE and PRE values parameter estimation under Breslow Method (ordering variable: sysbp)

		MSE and PRE values of Beta of Age		MSE and PRE values of Beta of sysbp	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,00511	**0,005082***	0,000263	**0,000213***
	PRE	100	**100,5468***	100	**123,1515***
30%	**MSE**	0,004207	0,004305	0,000309	**0,000197***
	PRE	100	97,71881	100	**156,8853***
60%	**MSE**	0,229129	**0,007376***	0,002092	**0,000301***
	PRE	100	**3106,421***	100	**694,7909***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	0,003181	0,000102	**0,000086***
	PRE	100	99,79307	100	**118,5415***
30%	**MSE**	0,001837	**0,001831***	0,000095	**0,000062***
	PRE	100	**100,3352***	100	**153,7798***
60%	**MSE**	0,001856	0,001891	0,000153	**0,000059***
	PRE	100	98,17256	100	**259,6574***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002842	**0,002821***	0,000069	**0,000062***
	PRE	100	**100,7291***	100	**110,646***
30%	**MSE**	0,001346	**0,001325***	0,000055	**0,000038***
	PRE	100	**101,5913***	100	**145,6014***
60%	**MSE**	0,000918	0,000925	0,000067	**0,000026***
	PRE	100	99,26825	100	**260,7236***
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002685	**0,002681***	0,000057	**0,000054***
	PRE	100	**100,1447***	100	**105,9282***
30%	**MSE**	0,001152	**0,001131***	0,000037	**0,000029***
	PRE	100	**101,807***	100	**125,0142***
60%	**MSE**	0,000511	**0,000507***	0,000032	**0,000013***
	PRE	100	**100,9037***	100	**247,214***

Table 8. MSE and PRE values of model under Breslow Method (ordering variable: sysbp)

Sample Size	**Censoring Rate**		**Sampling Method**	
			SRS	**RSS**
n=20	**%10**	**MSE**	0,002686	**0,002648***
		PRE	100	**101,46***
	%30	**MSE**	0,002258	**0,002251***
		PRE	100	**100,31***
	%60	**MSE**	0,115611	**0,003839***
		PRE	100	**3011,84***
n=50	**%10**	**MSE**	0,001638	**0,001633***
		PRE	100	**100,29***
	%30	**MSE**	0,000966	**0,000947***
		PRE	100	**102,09***
	%60	**MSE**	0,001005	**0,000975***
		PRE	100	**103,07***
n=100	**%10**	**MSE**	0,001455	**0,001442***
		PRE	100	**100,94***
	%30	**MSE**	0,000701	**0,000681***
		PRE	100	**102,81***
	%60	**MSE**	0,000492	**0,000475***
		PRE	100	**103,60***
n=200	**%10**	**MSE**	0,001371	**0,001367***
		PRE	100	**100,26***
	%30	**MSE**	0,000594	**0,000580***
		PRE	100	**102,39***
	%60	**MSE**	0,000272	**0,000260***
		PRE	100	**104,58***

Table 9. MSE and PRE values parameter estimation under Efron Method (ordering variable: sysbp)

		MSE and PRE values of Beta of Age		MSE and PRE values of Beta of sysbp	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,00511	**0,005083***	0,000263	**0,000214***
	PRE	100	**100,5392***	100	**123,078***
30%	**MSE**	0,004207	0,004308	0,000309	**0,000197***
	PRE	100	97,65614	100	**156,7893***
60%	**MSE**	0,229131	**0,007381***	0,002092	**0,000301***
	PRE	100	**3104,185***	100	**694,6371***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	0,003179	0,000102	**0,000086***
	PRE	100	99,8385	100	**118,5765***
30%	**MSE**	0,001837	**0,001831***	0,000095	**0,000062***
	PRE	100	**100,3547***	100	**153,7665***
60%	**MSE**	0,001857	0,001892	0,000153	**0,000059***
	PRE	100	98,14492	100	**259,5766***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002841	**0,002819***	0,000069	**0,000062***
	PRE	100	**100,7892***	100	**110,7073***
30%	**MSE**	0,001346	**0,001324***	0,000055	**0,000038***
	PRE	100	**101,6487***	100	**145,6679***
60%	**MSE**	0,000918	0,000925	0,000067	**0,000026***
	PRE	100	99,25262	100	**260,7205***
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002684	**0,002678***	0,000057	**0,000054***
	PRE	100	**100,234***	100	**106,0287***
30%	**MSE**	0,001152	**0,001130***	0,000037	**0,000029***
	PRE	100	**101,9034***	100	**125,1105***
60%	**MSE**	0,000511	**0,000507***	0,000032	**0,000013***
	PRE	100	**100,9126***	100	**247,2567***

Table 10. MSE and PRE values of model under Efron Method (ordering variable: sysbp)

Sample Size	Censoring Rate		Sampling Method	
			SRS	RSS
n=20	%10	MSE	0,002687	**0,002648***
		PRE	100	**101,45***
	%30	MSE	0,002258	**0,002253***
		PRE	100	**100,25***
	%60	MSE	0,115611	**0,003841***
		PRE	100	**3009,73***
n=50	%10	MSE	0,001638	**0,001633***
		PRE	100	**100,33***
	%30	MSE	0,000966	**0,000947***
		PRE	100	**102,10***
	%60	MSE	0,001005	**0,000975***
		PRE	100	**103,04***
n=100	%10	MSE	0,001455	**0,001441***
		PRE	100	**101,00***
	%30	MSE	0,000700	**0,000681***
		PRE	100	**102,87***
	%60	MSE	0,000492	**0,000475***
		PRE	100	**103,59***
n=200	%10	MSE	0,001370	**0,001366***
		PRE	100	**100,35***
	%30	MSE	0,000594	**0,000580***
		PRE	100	**102,49***
	%60	MSE	0,000272	**0,000260***
		PRE	100	**104,59***

Table 11. MSE and PRE values parameter estimation under Exact Method (ordering variable: sysbp)

		MSE and PRE values of Beta of Age		**MSE and PRE values of Beta of sysbp**	
n=20					
Censoring Rate		**Sampling Method**		**Sampling Method**	
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,005111	**0,005084***	0,000263	**0,000214***
	PRE	100	**100,5367***	100	**123,0196***
30%	**MSE**	0,004207	0,00431	0,000309	**0,000197***
	PRE	100	97,62632	100	**156,6882***
60%	**MSE**	0,248202	**0,007383***	0,00222	**0,000301***
	PRE	100	**3361,916***	100	**736,9444***
n=50					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,003174	0,003178	0,000102	**0,000086***
	PRE	100	99,85841	100	**118,5682***
30%	**MSE**	0,001837	**0,001830***	0,000095	**0,000062***
	PRE	100	**100,3767***	100	**153,7862***
60%	**MSE**	0,001857	0,001892	0,000153	**0,000059***
	PRE	100	98,13613	100	**259,5564***
n=100					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002841	**0,002817***	0,000069	**0,000062***
	PRE	100	**100,8491***	100	**110,7639***
30%	**MSE**	0,001346	**0,001323***	0,000055	**0,000038***
	PRE	100	**101,7157***	100	**145,7514***
60%	**MSE**	0,000918	0,000925	0,000067	**0,000026***
	PRE	100	99,27204	100	**260,7347***
n=200					
		SRS	**RSS**	**SRS**	**RSS**
10%	**MSE**	0,002684	**0,002676***	0,000057	**0,000053***
	PRE	100	**100,3023***	100	**106,1217***
30%	**MSE**	0,001151	**0,001129***	0,000037	**0,000029***
	PRE	100	**101,9925***	100	**125,2584***
60%	**MSE**	0,000511	**0,000507***	0,000032	**0,000013***
	PRE	100	**100,9533***	100	**247,4562***

Table 12. MSE and PRE values of model under Exact Method (ordering variable: sysbp)

Sample Size	Censoring Rate		Sampling Method	
			SRS	RSS
n=20	%10	MSE	0,002687	**0,002649***
		PRE	100	**101,44***
	%30	MSE	0,002258	**0,002254***
		PRE	100	**100,21***
	%60	MSE	0,125211	**0,003842***
		PRE	100	**3259,00***
n=50	%10	MSE	0,001638	**0,001632***
		PRE	100	**100,35***
	%30	MSE	0,000966	**0,000946***
		PRE	100	**102,13***
	%60	MSE	0,001005	**0,000975***
		PRE	100	**103,03***
n=100	%10	MSE	0,001455	**0,001440***
		PRE	100	**101,06***
	%30	MSE	0,000700	**0,000680***
		PRE	100	**102,93***
	%60	MSE	0,000492	**0,000475***
		PRE	100	**103,61***
n=200	%10	MSE	0,001370	**0,001365***
		PRE	100	**100,42***
	%30	MSE	0,000594	**0,000579***
		PRE	100	**102,58***
	%60	MSE	0,000272	**0,000260***
		PRE	100	**104,63***

References

[1] A.I. Al-Omari, *Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling*, Statist. Probab. Lett. **82**, 1883-1890, 2012.

[2] M.F. Al-Saleh and A.I. Al-Omari, *Multistage ranked set sampling*, J. Statist. Plann. Inference **102**, 273286, 2002.

[3] D. G. Altman, B.L. De Stavola, S.B. Love and K. A. Stepniewska, *Review of survival analyses published in cancer journals*, British Journal of Cancer **72**(2), 511, 1995.

[4] N. Ata Tutkun, N. Koyuncu and U. Karabey, *Discrete-time survival analysis under ranked set sampling: an application to Turkish motor insurance data*, J. Stat. Comput. Simul. **89**(4), 660-667, 2019.

[5] S.K. Ashour and M.S. Abdallah, *New distribution function estimators and tests of perfect ranking in concomitant-based ranked set sampling*, Comm. Statist. Simulation Comput. 1-26, 2019.

[6] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, *Survival analysis part II: multivariate data analysisan introduction to concepts and methods*, British Journal of Cancer **89**(3), 431-436, 2003.

[7] J. Borucka, *Methods of handling tied events in the Cox proportional hazard model*, Studia Oeconomica Posnaniensia **2**(2), 91-106, 2014.

[8] N.E. Breslow, *Covariance analysis of censored survival data*, Biometrics **30**, 89-99, 1974.

[9] H. Che, *Cutoff sample size estimation for survival data: a simulation study*, Unpublished Master Thesis, Department of Statistics, Uppsala University, Sweden, 2013.

[10] D. Collett, *Modelling survival data in medical research*, Chapman and Hall, UK, 1994.

[11] D.R. Cox, *Regression models and life tables (with discussion)*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **34**,187-220, 1972.

[12] B. Efron, *The efficiency of Coxs likelihood function for censored data*, J. Amer. Statist. Assoc. **76**, 312-319, 1977.

[13] J. Frey, *Nonparametric mean estimation using partially ordered sets*, Environmental and Ecological Statistic **19**, 309-326, 2012.

[14] N.M. Gemayel, E.A. Stasny, J.A. Tackett and D. A. Wolfe, *Ranked set sampling: An auditing application*. Review of Quantitative Finance and Accounting **39**, 413-422, 2012.

[15] F.Y. Hsieh and P. W. Lavori, *Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates*, Controlled Clinical Trials **21**,552560, 2000.

[16] A.A. Jemain and A.I. Al-Omari, *Multistage median ranked set samples for estimating the population mean*, Pakistan Journal of Statistics **22**,195207, 2006.

[17] A.A. Jemain, A.I. Al-Omari and K. Ibrahim, *Multistage extreme ranked set sampling for estimating the population mean*, J. Stat. Theory Appl. **6**(4),456471, 2007.

[18] J.D. Kalbfleisch and R.L. Prentice, *The statistical analysis of failure time data*, Wiley, New York, 1980.

[19] M. Mahdizadeh, and E. Zamanzade,*Smooth estimation of a reliability function in ranked set sampling*, Statistics **52**, 750-768, 2018.

[20] M. Mahdizadeh, and E. Zamanzade, *Interval estimation of $P(X < Y)$ in ranked set sampling*, Comput. Statist. **33**, 1325-1348, 2018.

[21] M. Mahdizadeh, and E. Zamanzade, *Efficient body fat estimation using multistage pair ranked set sampling*, Stat. Methods Med. Res. **28**: 223-234, 2019.

[22] M. G. Marmot, M. J. Shipley and G. Rose, *Inequalities in deathspecific explanations of a general pattern*, The Lancet **323**(8384), 1003-1006, 1984.

[23] G.A. McIntyre, *A method for unbiased selective sampling, using ranked sets*, Australian Journal of Agricultural Research **3**, 385390, 1952.

[24] M. Moerbeek, *Sufficient sample sizes for discrete-time survival analysis mixture models*, Structural Equation Modelling: A Multidisipliniary Journal **21**(1), 63-67, 2014.

[25] O. Ozturk, *Sampling from partially rank-ordered sets*, Environmental and Ecological Statistics **18**, 757-779, 2011.

[26] H. M. Samawi, A. Helu, H. Rochani, J. Yin, L. Yu and R. Vogel, *Reducing sample size needed for accelerated failure time model using more efficient sampling methods*, J. Stat. Theory Pract. **12**(3), 530-541, 2018.

[27] D. Schoenfeld, *Sample-size formula for the proportional-hazards regression model*, Biometrics **39**,499503, 1983.

[28] P. Royston, G. Ambler, and W. Sauerbrei, *The use of fractional polynomials to model continuous risk variables in epidemiology*, International Journal of Epidemiology, **28**(5), 964-974, 1999.

[29] S. Wang, J. Zhang and W. Lu, *Sample size calculation for the proportional hazards model with a time-dependent covariate*, Comput. Statist. Data Anal. **74**, 217-227, 2014.

[30] E. Zamanzade and M. Vock, *Variance estimation in ranked set sampling using a concomitant variable*, Statist. Probab. Lett. **105**, 1-5, 2015.

[31] E. Zamanzade and M. Mahdizadeh, *A more efficient proportion estimator in ranked set sampling* Statist. Probab. Lett. **129**, 28-33, 2017.

[32] E. Zamanzade and M. Mahdizadeh, *Distribution function estimation using concomitant-based ranked set sampling*, Hacet. J. Math. Stat. **47**(3), 755-761, 2018.

[33] E. Zamanzade and M. Mahdizadeh, *Estimating the population proportion in pair ranked set sampling with application to air quality monitoring*, J. Appl. Stat. **45**(3), 426-437, 2018.

[34] E. Zamanzade and M. Mahdizadeh, *Using ranked set sampling with extreme ranks in estimating the population proportion*, Stat. Methods Med. Res. **29** (1), 165-177, 2020.