Factors Affecting the Life Table Estimators

Marwa Omer Mohi-Eldin[1*] and Nidal Mohamed Mustafa[2]
[1]Department of Actuarial Sciences, Faculty of Mathematical Sciences and Statistics, Al-Neelain University, Khartoum, Sudan.
[2]Department of Statistics and Computation, Faculty of Mathematical Sciences and Statistics, Al-Neelain University, Khartoum, Sudan
[*]Corresponding author email: Memafo86@gmail.com

Abstract

This study is an attempt to know to what extent the efficiency of the estimates of the upper and lower quartiles of the life time functions, median life time and cumulative intensity rate of failure using life table method, are affected by factors such as; method of grouping data, censoring degree in the data, sample size and population distribution. The study is conducted and based on three generated exponentially distributed population with different parameters levels and along with different variances levels as well. The true values of interested parameters are computed first. Then estimates are computed from three sample sizes taken from each population. For each sample size three grouping levels and five censoring degrees are adopted. Mean Square Error (MSE) is used to compare the above mentioned property, and the inter quartile range is used as another measure of the efficiency. The results lead to the fact that the efficiency of the estimates increases with the increase in sample size for a given grouping and a given censoring degree, as well as the case of increasing with the increase in grouping levels for a given censoring degree and a given sample size. And the efficiency of the estimates decreases with the increase in censoring levels for a given sample size and a given grouping. For each sample size from each population, the study revealed that they have a positive skewedness. And for some cases of population three, precisely for small sample with grouping 8, and for big sample with grouping 12, they clearly tend to have a symmetric distribution.

Key words: Censoring, Grouping, Life Table, Intensity Rate of Failure, Population Distribution.

Introduction

The term life time data derives from the historical development of the field. John Graunt's (1662) book "Natural and political observations upon the bills of mortality", which classified registered deaths by age, period, gender and cause of death, suggested for the first time that death be regarded as an event which deserves systematic study. Some years later, Edmund Halley devised the first life table, very similar to those still in use today in demographic and actuarial studies, and Greenwood (1926) provided a variance formula for the life table estimator. Broadening the term survival analysis to include data on any event observed over time, not just death or failure (Akritas, 2004).

Life tables are one of the oldest statistical techniques and are extensively used by medical statisticians and by actuaries. It is the scheme for expressing the form of mortality in terms of probabilities. The life table is constructed from census data and death registration data, it can also be constructed from vital registration or retrospective surveys. Although the life table is one of the statistical tools most commonly used by applied statistics, rigorous derivations of many of its formal properties seem strangely to be lacking from the literature. For example, Greenwood's formula for the variance of the cumulative survival probability depends for its validity on the asymptotic independence of the estimates of the conditional probabilities of survival over the intervals used for grouping of the data. Chiang (1968) is often cited as a source for this results, although his proof applies only to the case of no live withdrawals. Breslow and Crowley (1974) outlined a general theory for the life table in which its familiar large sample properties. They established the consistency of standard (actuarial) life table under the large sample effect of censorship. Scherbov and Ediev (2011), studies the significance of life table estimates for small population, they considered the bias, standard error and distributions of the characteristics of life tables.

In this study the life table method is used to investigate the efficiency of the life time data parameters of the lower and upper quartiles, median survival time and cumulative force of mortality, to know to what extent the efficiency of theses parameters are affected by factors like the sample size, degree of censoring in the data, method of grouping data and population distribution. The study is conducted and based on three generated exponentially distributed population with different parameters levels as: 0.2, 0.8 and 1.2 respectively, and along with different variances levels as well. The true values of interested parameters are computed first. Then estimates are computed from three sample sizes taken from each population as: 200 as a big sample, 100 for medium sample and 50 for small sample. For each sample size three grouping levels 4, 8, and 12, and five censoring degrees, 3%, 5%, 7%, 10% and 30% are adopted. Thus the need arises to investigate, examine and to identify the magnitude of this effects. And this is the main problem which motivated this study.

Materials and Methods

This paper relies entirely on Life Table as the major method, along with Censoring, Grouping, cumulative Intensity Rate of Failure, Population Distribution, and Sample Size.

Life Table Method

Allison (1995), has observed that life table is among the most common methods used to analyze lifetime data. The primary function of life table is to summarize life time data grouped into intervals to provide estimates of the survivor function, the density function and the hazard rate. Kalbfleisch and Prentice (1980) and Allison (1984) have indicated that the life table is designed for situations where only the interval in which failure or censoring occurred is known but the actual failure or censoring time is unknown. Teachman (1983) has alluded that life tables are also useful for preliminary evaluation of data and evaluating the fit of regression models. It also allows assessment of exogenous variables in more complex analysis. It can also be used to assess the mortality level of a population and its age structure, project the population into the future, and assess the survival rate and the number of cases at risk within a cohort.

Construction of a Life Table

A life table is constructed from a set of grouped or ungrouped failure data. The columns of the table are created using a set of formulas, which will be defined bellow. The rows of the table represent different time intervals. [NCSS Statistical Software, NCSS.com]. see books of Lee(1992) and Elandt-Johnson and Johnson(1980)

1) Time interval

Each time interval is represented by $T_t \leq T < T_{t+1}$ or $[T_t, T_{t+1})$, where t=1, 2,…,s. The interval is from T_t up to but not includingT_{t+t}. The intervals are assumed to be fixed, and don't have to be of equal length, but it is often convenient to make them so. T_{mt} is a midpoint of the interval it is defined as halfway through the interval. The width of the interval is b_t where $b_t = T_{t+1} - T_t$. The width of the last interval b_s is theoretically infinite, so items requiring this value will be left of blank.

2) Number Lost to Follow-Up

The number lost to follow-up, c_i is the number of subjects that were loss to observation during this interval, so their survival status is unknown.

3) Number Withdrawn Alive

The number withdrawn alive, w_t , is the number of individuals who had not died (failed) by the end of the study.

4) Number Dying

It is a number of subjects that die (fail) during the interval , it's denoted by d_t

5) Number entering the t^th^ interval

In the first interval it is the total sample size, in the remaining interval it is computed using the formula:

$$\grave{n}_t = \acute{n}_{t-1} - c_{t-1} - w_{t-1} - d_{t-1}$$

6) Number Exposed to Risk

The number exposed to risk, n_t, is computed using the formula:

$$n_t = \acute{n}_{t-1} - \frac{1}{2}(c_{t-1} + w_{t-1})$$

This formula assumes that times to loss or withdrawal are distributed uniformly across the interval.

7) Conditional Proportion Dying

The conditional proportion dying, q_t, is an estimate of the conditional probability of death in the interval given exposure to the risk of death in the interval. It is computed using the formula:

$$q_t = \frac{d_t}{n_t}$$

8) Conditional proportion surviving

The conditional proportion surviving, p_t, is an estimate of the conditional probability of surviving through the interval. It is computed using the formula:

$$p_t = 1 - q_t$$

9) Cumulative proportion surviving

The cumulative proportion surviving, $S(T_t)$, is an estimate of cumulative survival rate at time T_t. It is computed using the formula:

$$S(T_t) = S(T_{t-1})p_{t-1}$$

Where; $S(T_1) = 1$

The variance of this estimate is itself estimated using the formula:

$$V[S(T_t)] = S(T_t)^2 \sum_{j=1}^{t-1} \frac{q_j}{n_j p_j}$$

10) Estimated Death Density Function:

The estimated death density function, $f(T_{mt})$, is an estimate of the probability of dying in the interval per unit width. At the interval midpoint it is computed using the formula:

$$f(T_{mt}) = \frac{S(T_t) - S(T_{t+1})}{b_t} = \frac{S(T_t)q_t}{b_t}$$

11) Hazard Rate Function

The estimated hazard rate function $h(T_{mt})$, is an estimate of the number of deaths per unit of time divided by the average number of survivors at the interval midpoint. It is computed using the formula:

$$h(T_{mt}) = \frac{f(T_{mt})}{S(T_{mt})}$$

$$= \frac{d_t}{b_t(n_t - \frac{1}{2}d_t)} = \frac{2q_t}{b_t(1+p_t}$$

12) Median Remaining Lifetime

The median remaining lifetime MRT_t, is the time value at which exactly one-half of those who survived until T_t are still alive. To compute this value find the value j such that $S(T_j) \geq \frac{1}{2}S(T_t)$ *and* $S(T_{j+1}) \leq \frac{1}{2}S(T_t)$. Next compute the median remaining lifetime using the formula:

$$MRT_t = (T_j - T_t) + \frac{b_j(S(T_j) - \frac{1}{2}S(T_t))}{S(T_j) - S(T_{j+1})}$$

Results

Combining the previous mentioned levels of factors, the results are presented in tables 1 to 12 below. Many of them are obtained, and many of them have the same results or approximately near to each other so they are not presented all. C_i is used to denote censoring degree, n_i is for sample size. Interpretation of every mentioned table is numbered and presented in a summarized way.

Population (1) - Exponential Distribution with parameter 0.2 and variance equal 25

Table 1: Estimates values of first quartile with its mean square error (between brackets) using life table method with grouping 4

C_i / n_i	0.03	0.05	0.07	0.1	0.3
50	0.0504 (0.0017)	0.0543 (0.0009)	0.0589 (0.0012)	0.0658 (0.0027)	0.1296 (0.0073)
100	0.0344 (0.0004)	0.0378 (0.0005)	0.04103 (0.0006)	0.0462 (0.0006)	0.1006 (0.0032)
200	0.0188 (0.0002)	0.0212 (0.0004)	0.0235 (0.0003)	0.0274 (0.0003)	0.0697 (0.0019)

Taking the sample size =50, we see that as the censoring degree increases, the estimates of first quartile increases. The same results are obtained with sample size =100 and 200.

Fixing the degree of censoring, we find that as the sample size increases the estimates of life time function and the mean square error decrease, which means that the estimates of life time function becomes efficient as the sample size increases with the same degree of censoring in the data.

The efficiency of the estimates decreases with the increase in censoring degree for a given sample size.

The most efficient estimate is at sample size of 200 and censoring level of 3%.

Table 2: Estimates values of second quartile (the median) with its mean square error using life table method with grouping 4

C_i / n_i	3%	5%	7%	10%	30%
50	0.1151 (0.0048)	0.1211 (0.0040)	0.1266 (0.0051)	0.1366 (0.0046)	0.2156 (0.0072)
100	0.0862 (0.0031)	0.0915 (0.0026)	0.0966 (0.0027)	0.1035 (0.0045)	0.1770 (0.0115)
200	0.0544 (0.0015)	0.0584 (0.0025)	0.0628 (0.0020)	0.0705 (0.0017)	0.1312 (0.0029)

The second quartile which is the median life time increases as the degree of censoring increases in the data for a given sample size and a given grouping.

The mean square error which measure the efficiency of the estimates explained that if 5% of the life time data (that has a sample size =50 or 100) was censored this may give the best estimates if the data can have a censoring degree less than 5% or more than 5%.

Given the degree of censoring, we see that as the sample size increases the median life time function decreases, and the efficiency of the estimates increases. The same result applies in all other factors levels combinations; there exists an increase on efficiency when the sample size increases and also a decrease in efficiency when the censoring degree increases. With some different points for some combinations for example: The results of the estimates of first quartile and its mean square error with grouping 12, at sample size=200, the efficiency of the estimates are stable, at censoring levels of: 3%, 5% and 7%. And the results of the estimates of the median life time function with its mean square error in grouping 12 shows that the most efficient estimates are at censoring degree of 5%, for all sample sizes.

Table 3: Estimates of inter quartile range with its mean square error using life table method with grouping 4

C_i / n_i	3%	5%	7%	10%	30%
50	0.3819 (0.0047)	0.3850 (0.0108)	0.3883 (0.0052)	0.3917 (0.0093)	0.4040 (0.0102)
100	0.3465 (0.0023)	0.3525 (0.0040)	0.3570 (0.0035)	0.3620 (0.0022)	0.3928 (0.0039)
200	0.3204 (0.0010)	0.3257 (0.0041)	0.3311 (0.0023)	0.3399 (0.0011)	0.3858 (0.0018)

From table 3 , it is clear that the estimates increase with increase in censoring degrees, and decrease with the increase in sample size. Which support the results of mean square error of the estimates.

For the rest two populations, and for all grouping levels and sample sizes, the same results as the table 3 are obtained.

Table 4: Estimates of the mean square error for cumulative intensity rate of failure using life table method with grouping 4

C_i / n_i	3%	5%	7%	10%	30%
50	0.0175 0.0008 0.7455 0.3087	2.9808 35.5493 16.8145 0.8070	2.8133 6.1505 29.2789 13.7986	0.0734 0.7114 7.1778 3.1432	0.3487 0.9178 1.7468 3.9708
100	0.8879 0.9601 0.0064 1.7518	0.3211 1.6481 0.7013 0.0021	1.3182 4.2344 28.3528 61.5823	0.7764 3.0749 1.7207 0.1107	0.3236 3.2611 0.8728 3.5654
200	0.0460 0.3324 0.0110 0.0221	0.1846 0.2732 1.1262 0.3962	0.4329 2.6674 3.9759 1.2999	0.0928 0.1928 0.0321 2.3128	0.0664 0.2701 0.7232 4.3025

At sample size =50, the efficient estimates are at censoring values: 3%, 10% and 30%.

At sample size =100, the efficient estimates are on censoring degrees 3%,5%,10%,30%.

At sample size =200, the efficient estimates are at all censoring degrees.

Estimates of the mean square error for intensity rate of failure, using life table method with grouping 8 and grouping 12, gives efficient estimates at all censoring degrees for all sample sizes.

Population (2) - Exponential Distribution with parameter 0.8, with variance equal 1.5625.

Table 5: Estimates of first quartile with its mean square error using life table method with grouping 4

C_i / n_i	0.03	0.05	0.07	0.1	0.3
50	0.1278 (0.0055)	0.1336 (0.0103)	0.1400 (0.0078)	0.1500 (0.0065)	0.2312 (0.0108)
100	0.0851 (0.0040)	0.0893 (0.0044)	0.0942 (0.0045)	0.1029 (0.0044)	0.1748 (0.0080)
200	0.0427 (0.0027)	0.0458 (0.0036)	0.0499 (0.0041)	0.0558 (0.0033)	0.1117 (0.0103)

At a specified sample size, the values of the estimates are increases with the increase in censoring degree. And the efficiency of the estimates decreases and increases.

At specified censoring degree, the estimates become more efficient as sample size increases.

When sample size=100, the same efficiency of the estimates are obtained with censoring degrees5%, 7% and 10%.

Below are some different results obtained from some combinations as:

- Estimates of median life time when grouping is 4 and sample size=100, the efficiency is extremely stable moving on censoring degrees.
- Estimates of first quartile with its mean square error using life table method with grouping 8 , the results obtained shows that at sample size =50, the efficiency of the estimates increases till censoring 7%. And for sample size =200, the most efficient estimates are at censoring degree 7%.

- Estimates of third quartile with its mean square error at grouping 12 shows that, when the sample size was100, the efficiency of the estimates are extremely stable.

Table 6: Estimates of mean square error for cumulative intensity rate of failure using life table method with grouping 4

C_i / n_i	0.03	0.05	0.07	0.1	0.3
50	129.2850 544.0941 402.2553 66.2134	0.4700 0.0355 0.2213 0.7647	116.1992 374.4532 373.1262 97.0300	107.2054 210.1537 288.7460 7.9768	61.1645 131.1859 121.3094 35.8661
100	45.3725 406.4812 12.3277 79.9598	18.2593 101.6577 24.5269 52.7697	5.0396 363.6624 22.7560 50.9628	104.8778 171.4106 2.2631 162.7959	30.9421 93.4979 19.1072 18.4710
200	0.0871 34.6155 161.7354 237.0984	0.0207 0.3450 0.2585 0.1665	0.3116 1.2243 2.5064 1.6193	0.3225 0.5746 0.1535 0.0094	0.0001 0.0975 0.4369 1.2622

The efficiency of estimates is clear at sample size =50, at censoring 5%. While, at sample size =200, it is also clear but at censoring degrees 5%,7%,10% and 30%.

Table 7: Estimates of mean square error for cumulative intensity rate of failure using life table method with grouping 8

C_i / n_i	3%	5%	7%	10%	30%
50	0.0013 0.0183 0.0245 0.0886 1.0287 0.7686 0.7378 0.7367	1.1273 1.8253 11.8600 53.4926 7.9857 2.6866 4.8233 4.8725	1.3167 1.3008 0.1806 0.5227 27.1674 93.2282 105.2763 105.8003	1.1484 16.7850 41.6037 12.0822 0.3860 23.8392 18.8679 18.7212	0.0006 0.0001 0.1567 0.9420 1.2076 0.9936 0.9666 0.9653
100	6.1214 21.7215 36.0822 38.1475 22.5915 1.1827 13.5343 15.1501	4.9952 42.8534 0.8399 390.0376 123.1062 4.8804 0.5233 0.7475	0.0051 0.0481 0.0017 0.0003 0.1738 0.4719 0.5868 0.5955	0.0049 0.0014 0.2865 0.0215 0.7664 1.7068 1.4941 1.4798	0.1294 0.3149 0.9067 1.0180 0.3729 0.1692 0.1213 0.1183
200	0.0245 0.1219 0.5246 0.4219 0.1736 0.0799 0.0470 0.0425	0.0285 0.1833 0.4515 0.3640 0.1402 0.0592 0.0320 0.0283	21.1474 105.7431 157.0505 66.0913 57.7210 62.9813 3.4000 1.0320	0.0189 0.1073 0.1789 0.1491 0.0010 0.0324 0.4648 0.4194	0.0113 0.0263 0.1025 0.0554 0.3813 0.8322 1.1332 1.1881

Regarding the sample size of 50, one can realize the efficiency of estimates for the censoring degree 3% and 30%

For sample =100, the efficiency of estimates is at censoring degrees 7%, 10% and 30%.

Unfortunately, the only inefficient estimates for sample =200 are at censoring 7%.

Table 8: Estimates of mean square error forcumulative intensity rate of failure using life table method with grouping 12

C_i / n_i	3%	5%	7%	10%	30%
50	0.0000 0.0009 0.0087 0.8326 1.4264 1.0765 1.0261 1.0227 1.0227 1.0227 1.0227 1.0227	0.0000 0.0080 0.0077 0.3253 0.9420 1.2614 1.3196 1.3232 1.3233 1.3233 1.3233 1.3233	0.0000 0.0486 0.0359 0.0002 0.1485 0.2850 0.3131 0.3147 0.3148 0.3148 0.3148 0.3148	0.0000 0.0261 0.0339 0.3198 0.7786 0.5403 0.5026 0.5001 0.5001 0.5001 0.5001 0.5001	0.0000 22.0183 140.6791 27.1527 41.5379 617.6204 589.1703 586.7152\ 586.6081 586.6081 586.6081 586.6081
100	0.0000 0.0057 0.0183 0.0074 0.1394 0.4965 0.6760 0.7017 0.7036 0.7036 0.7036 0.7036	0.0000 0.0022 0.0055 0.0154 0.1065 0.4212 0.5835 0.6078 0.6094 0.6094 0.6094 0.6094	0.0000 0.8594 2.8874 50.1122 3.8914 24.1881 184.9611 173.1395 172.4854 172.4854 172.4854 172.4854	0.0000 0.0098 0.0120 0.0325 0.4689 1.1588 0.9350 0.9040 0.9030 0.9030 0.9030 0.9030	0.0000 0.0016 0.0030 0.0194 0.1053 0.2462 1.2183 1.1975 1.1963 1.1963 1.1963 1.1963
200	0.0000 0.0002 0.0005 0.0038 0.0064 0.2396 0.6276 0.7699 0.7868 0.7868 0.7868 0.7868	0.0000 0.0041 0.0102 0.0069 0.0003 0.3205 0.7362 0.8961 0.9158 0.9158 0.9158 0.9158	0.0000 0.0006 0.0001 0.0055 0.0283 0.0988 0.5096 0.3985 0.3880 0.3880 0.3880 0.3880	0.0000 22.1141 32.9043 99.6983 0.9134 4.4558 47.4776 90.5223 95.7835 95.7835 95.7835 95.7835	0.0000 1.6470 0.0041 2.6548 47.2942 186.7718 379.7119 302.0041 294.3038 294.3038 294.3038 294.3038

From table number (8) and at censoring degrees of 3% and 5%, the estimates of hazard are efficient for the three sample sizes.

At censoring 7%, the efficient estimates are at sample sizes of 50 and 200.

At censoring 10%, the efficient estimates are at sample sizes of 50 and100.

At censoring degree 30%, the efficient estimates are only at sample size of 100. **Population (3) - Exponential with parameter 1.2, and variance 0.694444.**

Table 9: Results of first quartile with its mean square error using life table method with grouping 8

C_i / n_i	3%	5%	7%	10%	30%
50	0.1580 (0.0170)	0.1644 (0.0202)	0.1658 (0.0235)	0.1823 (0.0371)	0.2623 (0.0226)

100	0.0891 (0.0503)	0.0913 (0.0079)	0.0982 (0.0102)	0.1068 (0.0178)	0.1775 (0.0147)
200	0.0391 (0.0030)	0.0420 (0.0030)	0.0453 (0.0028)	0.0497 (0.0043)	0.0993 (0.0074)

As table 9 shows, the estimates become more efficient with the increase in sample size. But this is not applic able for the case of 3% of censoring; it first decreases a nd then increases. The same results are reached for est imates of median life time table, and third quartile.

The results of median life time with its mean square er ror using life table method with grouping4: this results shows that at a given censoring degree, the efficiency o f the estimates increases at sample size of 100 and then decrease at sample size of 200.

Table 10: First quartile with its mean square error using life table method with grouping 12

n_i \ C_i	3%	5%	7%	10%	30%
50	0.1551 (0.0211)	0.1621 (0.0407)	0.1699 (0.0193)	0.1796 (0.0368)	0.2603 (0.0382)
100	0.0896 (0.0083)	0.0942 (0.0128)	0.0988 (0.0091)	0.1067 (0.0160)	0.1744 (0.0299)
200	0.0389 (0.0034)	0.0412 (0.0029)	0.0451 (0.0027)	0.0491 (0.0045)	0.0991 (0.0083)

In table 10, the estimated values of the first quartile ar e exactly the same as the estimated values of the medi an life time.

At sample =200, the efficiency of the estimates increa se with the increase in censor degrees from3% to 7%.

Table 11: Mean square error for cumulative intensity rate of failure using life table method with grouping4

n_i \ C_i	3%	5%	7%	10%	30%
50	10.8229 14.5675 393.5047 674.1695	599.6041 20.7439 303.2443 545.7046	10.7849 0.1511 62.7155 62.2526	9.0635 0.8509 198.6222 391.6080	5.3379 8.1011 46.9210 6.3821
100	45.1060 76.3866 16.4824 0.4458	42.2911 200.0713 16.5382 1.7218	38.5864 55.5819 17.0823 2.4355	36.7096 52.4158 3.3399 5.7656	21.3970 40.8907 7.6419 81.7078
200	0.1604 0.7761 0.3436 0.0632	0.1524 0.7415 0.0769 0.4279	0.1432 0.7119 0.6340 1.1057	0.0332 0.5400 0.4311 0.2666	71.2897 309.5625 72.8149 72.3192

From the above results we know that the efficient estim ates are at sample =200 with censoring 3%, 5%, 7% an d 10%.

Table 12: Results of cumulative intensity rate of failure with its mean square error using life table method with g rouping8:

n_i \ C_i	3%	5%	7%	10%	30%
50	0.0000 0.1008 0.0346 0.1854 0.2403 0.2451	0.0000 0.0045 0.1297 0.3585 0.4337 0.4387	0.0000 2.9265 159.8394 142.3048 98.2771 95.6953	0.0000 0.0978 0.2253 1.3616 1.2347 1.2257	0.0000 86.1689 0.4717 35.0418 53.0650 54.5331

	0.2452 0.2452	0.4388 0.4388	95.6057 95.6057	1.2255 1.2255	54.5331 54.5331
100	0.0000 0.0001 0.1052 0.2211 0.2732 0.2816 0.2820 0.2820	0.0000 0.1666 82.0863 1.7137 39.9890 51.2425 51.6893 51.6893	0.0000 138.5960 208.1658 119.4212 33.8010 25.4405 25.1392 25.1392	0.0000 0.0436 0.0553 1.2681 2.7592 2.6839 2.6794 2.6794	0.0000 2.5451 28.4427 1.5108 8.5270 12.3773 12.6559 12.6559
200	0.0000 0.0024 0.0002 0.2525 0.3348 0.2227 0.2119 0.2119	0.0000 0.0024 0.0549 0.2959 0.7463 0.9410 0.9649 0.9649	0.0000 0.0001 0.1333 0.1191 0.4213 0.5619 0.5803 0.5803	0.0000 0.0220 0.0056 0.6062 1.3605 1.1378 1.1152 1.1152	0.0000 1.5358 112.6441 89.7720 84.6419 46.1326 42.2534 42.2534

At sample =50 the efficient estimates are at censoring 3 %,5% and 10%. And as well as at censoring degree = 7 % for estimated hazard with grouping 12.

At sample =100, the efficient estimates are when censo ring was 3% and 10%.

At sample =200, the efficient estimates are when censo ring 3%,5%,7%,and 10%. This is true when grouping w as 12 also, except for censoring 7% of the data.

Demonstration by means of figures

Below the results of figures: also at most combinations the same figures are obtained, so a few of them are pre sented .

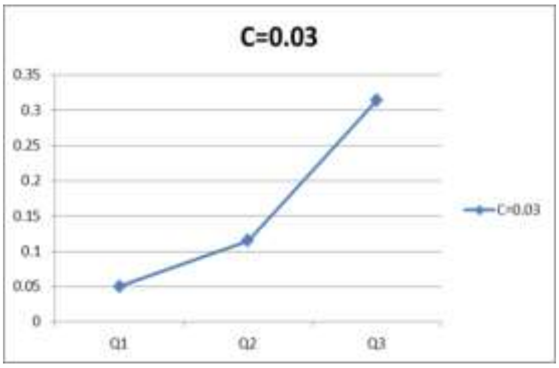

Figure 1: Estimates of quartiles,Q_1:lower, Q_2median,Q_3upper, opulation1

Figure 1 shows the estimated quartiles using life table method with grouping 4 and sample size 50,censoring degree =3%.

From the graph it is obvious that the median is near to first quartile what's mean that the distribution is positively skewed.

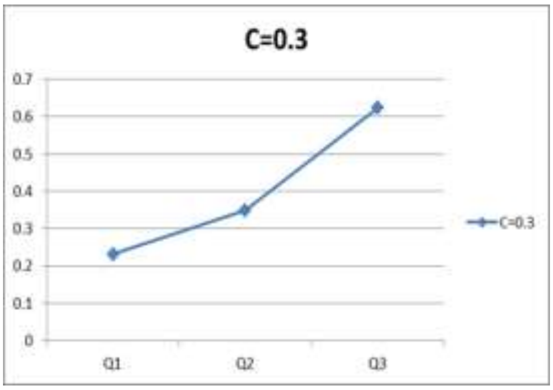

Figure 2: Estimates of quartiles,Q_1:lower, Q_2median, Q_3upper, opulation2

Figure 2 shows the estimates of quartiles by using life table method sample size =50, and grouping=4,censor sing=30% of the data of popultion2.

Also the median life time is near to the first quartile, w hich means that the distribution is positively skewed.

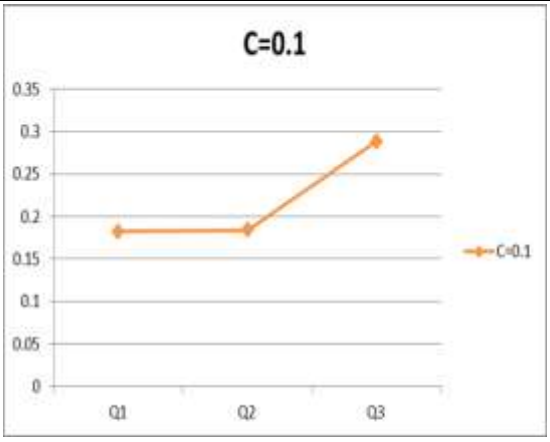

Figure 3: Estimates of quartiles,Q_1:lower, Q_2median ,Q_3upper, opulation3

Figure 3 is about the estimates of quartiles by using lif e table method sample size=50, and grouping = 8,cens oring =10% of the data from popultion3.

In this figure the first quartile is close to the median lif e time.

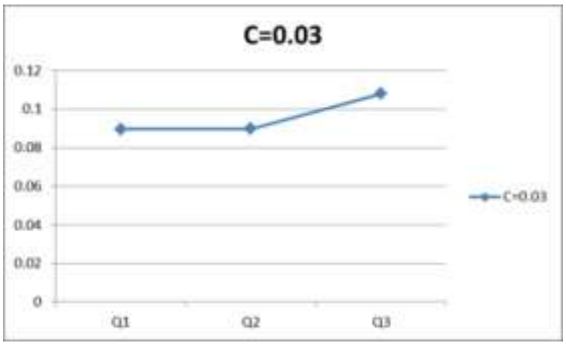

Figure 4: Estimates of quartiles,Q_1:lower, Q_2median ,Q_3upper, opulation3

Figure 4 shows the estimates of quartiles by using life table method sample size =100, and grouping =12,cen soring =3% of the data of popultion3.

It seems that the distances are approximately equal bet ween the median and third quartile, so the distribution in this case may be a symmetric.

Conclusion

The estimates of the first quartile of the life time distribution function, median life time and the third quartile of the life time functions which were obtained for the three populations, are proved to become more efficient by the increase in sample sizes. This result is consistent with what Scherbov and Ediev (2011) obtained on their paper "Significance of life table estimates for small populations". The efficiency of the estimates for the three populations increases, decreases and then increases, or remain the same for a few combinations, or sometimes it tends to increase with the increase in censoring degree for a given sample size.

If comparisons are made with respect to grouping levels, the efficiency differs from population to another. For example in population one the estimates of median life time are more efficient in grouping 4 and 12, whereas, at population two the efficient estimates of median life time are obtained when grouping level was 12 at the biggest sample size and so on. There is a little decrease in the efficiency of the estimates obtained for population two than one, while the least efficiency level is at population three, i.e. the efficiency of the estimates decreases with decrease in variance of the population.

The results of the hazard functions show that, the efficiency of the hazard function decreases with the decrease in population variance. The most inefficient estimates of hazard rates are at population two when grouping was 4. It seems that the previous two points comes with what is not familiar in the literature that the efficiency of the estimates increases with the decrease in variance. And this will need more research so as to be confirmed or disconfirmed.

The distribution of samples from population one and two are positively skewed as the above figures show. In population two with the small sample and grouping12 for all censoring degrees, the values of first quartile and median life time are close to them. This is also true in population three with grouping 8 for small and medium sample sizes. The distribution of samples from population three is positively skewed as the above figures show. Except, at grouping 12 with medium sample size it seems to be an asymmetric distribution.

References:

Allison, P. D. (1995).Survival Analysis Using the SAS System: A Practical Guide. SAS institute, Inc., Cary

Allison, P. D. (1984). Event History Analysis. Sage Publications, Beverly Hills.

Akritas, M.G.(2004). Nonparametric Survival Analysis: institute of Mathematical Statistics.

Breslow, N., Crowley, J. (1974)., A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2,437-453.

Chiang, C.L. (1968).Introduction to Stochastic Process in Biostatistics. Wiley, New York.

Graunt, J. (1662). Natural and political Observations made upon the Bills of mortality .London.

Greenwood, M. .(1926). The natural duration of cancr. Reports in public Health and medical Subjects33.H.M. Stationary office, London.

Kalbfleisch, J.D. (1980). and Prentice, R.L The Statistical Analysis of Failure Time Data. John Wiley and sons, New York..

Sergei S. Ediev D. (2011)., “significance of life table estimates for small populations: simulation-based study of standard errors”. Max Plank Institute for Demographic Research.22,527-550.

Teachman, J.D. (1983). Analyzing Social Process: life table and proportional hazards model. Social science research, 12,263-301.