DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ESTIMATION OF HARMONIC AND GEOMETRIC MEAN IN RANKED SET SAMPLING

by
Sukma Adi PERDANA

October, 2014
İZMİR

ESTIMATION OF HARMONIC AND GEOMETRIC MEAN IN RANKED SET SAMPLING

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Statistics

by
Sukma Adi PERDANA

October, 2014
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"ESTIMATION OF HARMONIC AND GEOMETRIC MEAN IN RANKED SET SAMPLING"** completed by **SUKMA ADI PERDANA** under supervision of **ASSIST. PROF. DR. NESLİHAN DEMİREL** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Neslihan DEMİREL

Supervisor

Assoc. Prof. Dr. Özlem EGE ORUÇ

(Jury Member)

Assoc. Prof. Dr. Engin MERMUT

(Jury Member)

Prof. Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude for my advisor Assist. Prof. Dr. Neslihan DEMİREL, for her support and encouragement that she gave to me so I have accomplished my research and thesis study.

I would like to thank all of academic staff in Statistics Department especially to Assoc. Prof. Dr. Özlem EGE ORUÇ and Assoc. Prof. Dr. Selma GÜRLER for their advice and support.

I want to say my big gratitude to my family especially for my mother Mulyani and my father Sukroni who had given the best support to me. I wish to express my gratitude also for my colleagues for their help and my PPI Izmir especially for Bayu meta anggoro. Last, I want to express my biggest gratitude for Republic of Turkey especially for YTB that had been supporting me and giving a broad opportunity for me so that I can study in Turkey.

Sukma Adi PERDANA

ESTIMATION OF HARMONIC AND GEOMETRIC MEAN IN RANKED SET SAMPLING

ABSTRACT

In this study, we develop an estimator for harmonic and geometric mean of population in ranked set sampling design. Ranked set sampling is a set of sampling units which is drawn from population can be ranked by using another certain mean without the actual measurement of the variable of interest which is to measure this variable is more costly and time consuming. The harmonic and geometric mean have numerous engineering applications, for instance: design stream flow estimation for waste load application areas, or estimation of effective petrochemical and geophysical properties of a heterogeneous system of porous media. The study investigates the bias and relative efficiency of the proposed mean estimators. The efficiency comparison is made for some distributions. It is shown that the new estimators out performs its competitors in the literature.

Keywords: Ranked set sampling, simple random sampling, harmonic mean, geometric mean, mean estimator, relative efficiency.

SIRALI KÜME ÖRNEKLEMESİNDE HARMONİK VE GEOMETRİK ORTALAMA KESTİRİMİ

ÖZ

Bu çalışmada, sıra sayılı küme örneklemesi deseninde kitle harmonik ve geometrik ortalamaları için kestirici geliştirilmiştir. Sıra sayılı küme örneklemesi; ilgilenilen değişken için gerçek ölçüm yapılmadan, ilişkili bir başka değişkenin değerine göre sıralayarak, kitleden çekilen örneklem birimlerinin kümesini veren, zamandan ve maliyetten tasarruf sağlayan bir örnekleme yöntemidir. Harmonik ve geometrik ortalamanın mühendislikte birçok uygulama alanı mevcuttur, örnek olarak atık yükü uygulama alanları için akarsu akış tahmini tasarımı ve gözenekli ortamın heterojen bir sistemin petrokimya ve jeofizik özelliklerinin etkinlik tahmini verilebilir. Bu çalışmada önerilen ortalama kestiricileri için yansızlık ve göreli etkinlikler araştırılmıştır. Etkinlik karşılaştırmaları birkaç farklı dağlım için yapılmıştır. Önerilen kestiricilerin literatürde var olan rakipleri ile yarışacak düzeyde iyi sonuçlar verdiği gösterilmiştir.

Anahtar kelime: Sıra sayılı küme örneklemesi, basit rastgele örnekleme, harmonik ortalama, geometrik ortalama, ortalama kestiricisi, göreli etkinlik.

CONTENTS

Page

M.Sc THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES viii
LIST OF TABLES ix

CHAPTER ONE-INTRODUCTION 1

CHAPTER TWO-SIMPLE RANDOM SAMPLING AND RANKED SET SAMPLING.....3

2.1 Simple Random Sampling 3
2.2 Ranked Set Sampling.....5

CHAPTER THREE-MEAN ESTIMATOR OF HARMONIC MEAN AND GEOMETRIC MEAN BASED ON RANKED SET SAMPLING 13

3.1 Geometric Mean 13
3.2 Harmonic Mean 14
3.3 Proposed Estimator HM and GM in Ranked Set Sampling.....15
3.3.1 Proposed Estimator of HM 15
3.3.2 Proposed Estimator of GM 15

CHAPTER FOUR-SIMULATION STUDY 16

4.1 Simulation Process 16
4.2 Simulation Result 20

4.2.1 Simulation Result of Simulation Process by Using The Correlation Coefficient Model of Dell and Clutter with Modification Based on Regression 20
4.4.2 Simulation Result of Simulation Process by Using The Correlation Coefficient from The Imperfect Judgment Ranking Model of Dell & Clutter and Adam & Levine 29

CHAPTER FIVE-CONCLUSION 37

REFERENCES 38

APPENDIX 40

LIST OF FIGURES

Page

Figure 2.1 How RSS can span the full range of the population 3
Figure 2.2 RSS procedure 7
Figure 4.1 Histogram of generating data from different distributions 17
Figure 4.2 Graphics of comparison of relative efficiency for normal distribution.....23
Figure 4.3 Graphics of comparison of relative efficiency for lognormal distribution 24
Figure 4.4 Graphics of comparison of relative efficiency for gamma distribution 25
Figure 4.5 Graphics of comparison of relative efficiency for exponential distribution 26
Figure 4.6 Graphics of comparison of relative efficiency for normal distribution from simulation process by using original Dell & Clutter model 31
Figure 4.7 Graphics of comparison of relative efficiency for lognormal distribution from simulation process by using original Dell & Clutter model 32
Figure 4.8 Graphics of comparison of relative efficiency for gamma distribution from simulation process by using original Dell & Clutter model 33
Figure 4.9 Graphics of comparison of relative efficiency for exponential distribution from simulation process by using original Dell & Clutter model 34

LIST OF TABLES

Page

Table 4.1 Relative Efficiencies (RE) of AM, HM, and GM for normal distribution .23

Table 4.2 Relative Efficiencies (RE) of AM, HM, and GM for lognormal distribution.....24

Table 4.3 Relative Efficiencies (RE) of AM, HM, and GM for gamma distribution.....25

Table 4.4 Relative Efficiencies (RE) of AM, HM, and GM for exponential distribution.....26

Table 4.5 Biases of AM, GM, and HM for normal and exponential distribution 27

Table 4.6 Biases of AM, GM, and HM for gamma and lognormal distribution 28

Table 4.7 Relative Efficiencies (RE) of AM, HM, and GM for normal distribution from simulation process by using original Dell & Clutter model 31

Table 4.8 Relative Efficiencies (RE) of AM, HM, and GM for lognormal distribution from simulation process by using original Dell & Clutter model 32

Table 4.9 Relative Efficiencies (RE) of AM, HM, and GM for gamma distribution from simulation process by using original Dell & Clutter model 33

Table 4.10 Relative Efficiencies (RE) of AM, HM, and GM for exponential distribution from simulation process by using original Dell & Clutter model 34

Table 4.11 Biases of AM, GM, and HM for normal and exponential distribution from simulation process by using original Dell & Clutter model 35

Table 4.12 Biases of AM, GM, and HM for gamma and lognormal distribution from simulation process by using original Dell & Clutter model 36

CHAPTER ONE

INTRODUCTION

One of the important points in statistics is the principle to guarantee that sample data properly represents the underlying population. The satisfaction of this principle provides the basis upon which statistical inference can be made for sample data. Sample data is the subset of population which is drawn in order to estimate the parameter of population. Simple random sampling (SRS) is generally used to draw sample data. However, this method relatively does not cover the full range of population. This is important because it can reduce the accuracy of estimator of the parameter of population. Alternatively, there is better method which can be done to take the sample data more accurately which is by using ranked set sampling (RSS).

The goal of ranked set sampling is to obtain more representative sample of population by collecting observations from a population that are more likely to span the full range of values the population (Wolfe, 2012). Dell & Clutter (1972) stated that the RSS is the most useful sampling technique to drawn sampling when measurement of the object sampling is difficult but it can be sampled and ranked by a judgment or concomitant variable. By using relative precession (RP), Dell & Clutter (1972) also showed that efficiency of the mean by RSS is better than SRS for estimating mean. Dell & Clutter (1972) established that the balanced RSS estimator for mean is unbiased estimator for the population mean regardless whether our judgment ranking is perfect or imperfect. The population mean, which is discussed by Dell & Clutter (1972) is arithmetic mean whereas there are two others mean which can also be used as central tendency: harmonic and geometric mean.

Harmonic mean is one of the measurements that can be used as a central tendency of a random variable in engineering applications. For instance, estimation of the design stream flow for waste load allocations (Rossman,1990), estimation of the effective or aggregate properties of geophysical media (Gomez-Hernandez and Gorelick ,1989) and the design of declining rate filter beds (Saatci ,1989). Geometric mean is also one of the measurements that can be used as a central tendency of random variables in real life applications. For instance, estimating the future growth

rate in economics (Spizman, 2008), report fecal coliform bacteria concentrations for water quality standards, and for evaluate investment returns and fluctuating interest rates. Nevertheless, based on our knowledge there is no literature about estimating harmonic mean and geometric mean in ranked set sampling.

The purpose of this study is to propose the mean estimators for the population harmonic and geometric mean in ranked set sampling design. The proposed estimators are evaluated by investigates the bias and relative efficiency of the proposed mean estimators for some distributions.

The study consists of five chapters. The basic concept of ranked set sampling (RSS) is described in second chapter. In this chapter, processes of ranking and theoretical proof of estimator are also presented. In third chapter, mean estimator of harmonic mean and geometric mean based on RSS design are discussed. In fourth chapter, simulation study is done by generating bivariate distributions and evaluating relative efficiency based on ratio of SRS and RSS. In order to measure how perfect is the relation between bivariate data (X,Y), two different correlation coefficient were used. One of the correlation coefficient is from the imperfect judgment ranking model of Dell & Clutter (1972) and Adam & Levine (1972). The other is the correlation coefficient model of Dell & Clutter (1972) with modification based on regression concept. In last chapter, result of simulation study and conclusion of this study is summarized.

CHAPTER TWO

SIMPLE RANDOM SAMPLING AND RANKED SET SAMPLING

2.1 Simple Random Sampling

Simple random sampling is the basic sampling technique in the literature. A sample design that used for obtaining n units sample data from N unit population data so that each combination $\binom{N}{n}$ of certain sample data has an equal probability to be drawn is simple random sampling (Cochran, 1977). Wolfe (2012) gives the definition that simple random sampling (SRS) of size n from an underlying probability distribution with probability density function $f(x)$ and cumulative probability function $F(x)$ is a collection of random variables $X_1, \dots, X_n$ which satisfy the following two properties:

(i) Each $X_i, i = 1, \dots, n$ has the same probability distribution.

(ii) The n random variables $X_1, \dots, X_n$ are mutually independent.

In order to use SRS, a list of all observation units in the population is needful. Observation unit is the same with the sampling unit. The sampling unit is assigned with a number. When the sample is drawn, each sampling unit has equal probability to be occurred in the sample and the selection of sampling unit is not affected by the sampling unit that has been selected previously. If there are N units population and its n units are taken as the sample then the probability to take the sample data of n units is calculated by the following equation (Lohr, 1999).

$$P(S) = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!} \tag{2.1}$$

Simple random sampling without replacement is more commonly used than SRS with replacement, because SRS without replacement is more efficient. In order to select the sample by SRS without replacement, there are some methods that can be utilized for instance by using tables of random number or by using computer-generated pseudorandom numbers.

One of purpose of survey sampling is doing inferences about population parameter by using information from a selected sample. The objective of this inference is often to estimate the population mean, especially the arithmetic mean (AM), denoted by μ and can be estimate by $\bar{x}$ as Equation 2.2.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.2}$$

The variance of $\bar{x}$ can be calculated by the following equations:

$$V(\overline{X}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) \tag{2.3}$$

$$V(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \tag{2.4}$$

Where $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}$ is the sample variance (2.5)

It can be shown that $\bar{x}$ is an unbiased estimator of μ by explanation as this below

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\, n\, \mu = \mu$$

And also it can be shown that $\frac{s^2}{n}\left(\frac{N-n}{N}\right)$ is an unbiased estimator of $V(\bar{x})$ by utilizing Equation 2.4 and the explanation is as follows (Scheaffer, 1990, p. 378).

$$E(s^2) = E\left[\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]$$

$$= \left(\frac{1}{n-1}\right)E\left[\sum_{i=1}^{n}\left[\left(X_i-\mu\right)-\left(\overline{X}-\mu\right)\right]^2\right]$$

$$= \left(\frac{1}{n-1}\right)E\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2 - n\left(\overline{X}-\mu\right)^2\right]$$

$$= \left(\frac{1}{n-1}\right)\left[\sum_{i=1}^{n}E\left(X_i-\mu\right)^2 - nE\left(\overline{X}-\mu\right)^2\right]$$

$$E(s^2)=\frac{1}{n-1}\left[n\sigma^2-nV\left(\bar{X}\right)\right]=\frac{1}{n-1}\left[n\sigma^2-n\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\right]$$

$$=\frac{\sigma^2}{n-1}\left(n-\frac{N-n}{N-1}\right)=\frac{N}{N-1}\sigma^2$$

Therefore by Equation 2.4

$$E\left[\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)\right]=\left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{N}{N-1}\right)\sigma^2=\left(\frac{N-n}{N-1}\right)\left(\frac{\sigma^2}{n}\right)=V\left(\bar{X}\right)$$

$\left(\frac{N-n}{N}\right)$ is finite population factor, which decrease the variance of estimator when the sample size n is large relative to population size N. Generally, the finite population factor can be ignored if $(N-n)/N \geq 0.95$ or equivalently $\frac{n}{N} \leq 0.05$ (Scheaffer, 1990).

2.2 Ranked Set Sampling

Ranked set sampling is proposed by McIntyre for the first time in 1952 in order to estimate the yield of pasture in Australia. The basic idea of ranked set sampling (RSS) is a set of sampling units which are drawn from population that can be ranked by using another certain mean without the actual measurement of the variable of interest when to measure this variable is more costly and time consuming (Chen, 2003). In other words, selecting a set from finite or infinite population by ranked data and using this data for inferences is easier and cheaper than traditional sampling methods. The purpose of RSS is to cover the full range of the population as illustrated in Figure 2.1. In this figure, the sample data which is obtained by SRS design is indicated by the curve with the full line, whereas the curve with dotted line indicates sample data which is taken by RSS design. As we can see in Figure 2.1, the sample data with RSS design shows having wider range that cover the full range of the population. Therefore, the estimator of parameter can provide better information about the parameter.

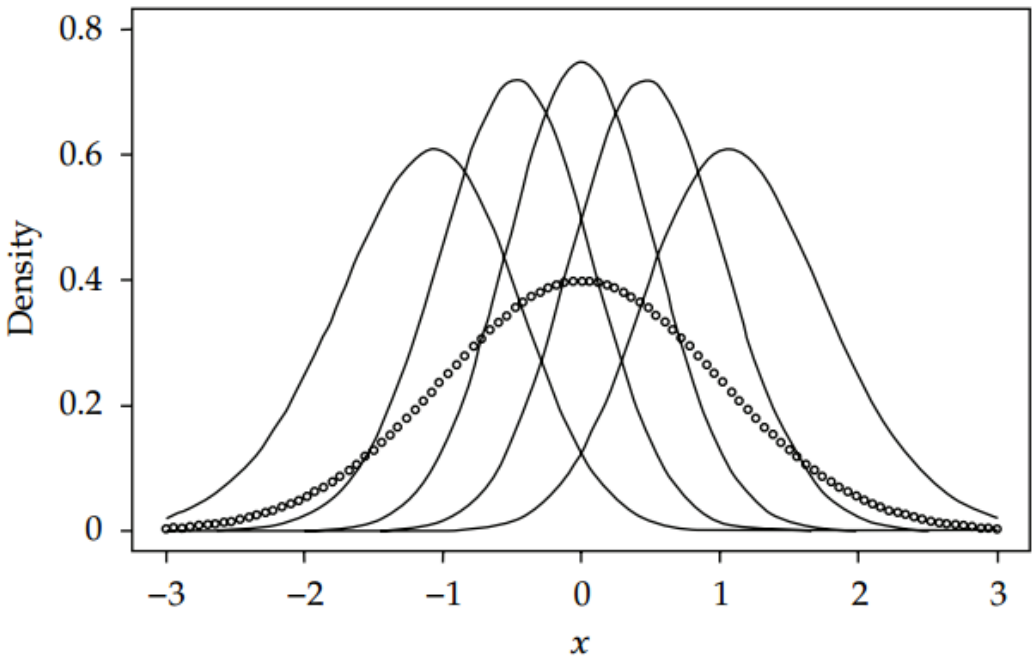


Figure 2.1 How RSS can span the full range of the population (Wolfe, 2012)

From (Chen, Bai, and Sinha, 2003), the original form of RSS which is proposed by McIntyre can be described as follows. First, selecting sample of size h by simple random sample from the underlying population and h sampling unit are ranked with respect to the concomitant variable of X, without actual measurement of X. and then, the first rank of this unit is identified and taken for the measurement of X. The remaining units of the sample are discarded. Furthermore, another same process is done and again the units of the sample are ranked by concomitant variable of X, the second rank unit is taken for the measurement of X and the remaining units are discarded. This process is done as many as h times and it can be obtained h set unit. This step is said to as a cycle. The cycle is repeated as much as c times and yields RSS of size n = c h. The sampling procedure can be illustrated in Figure 2.2.

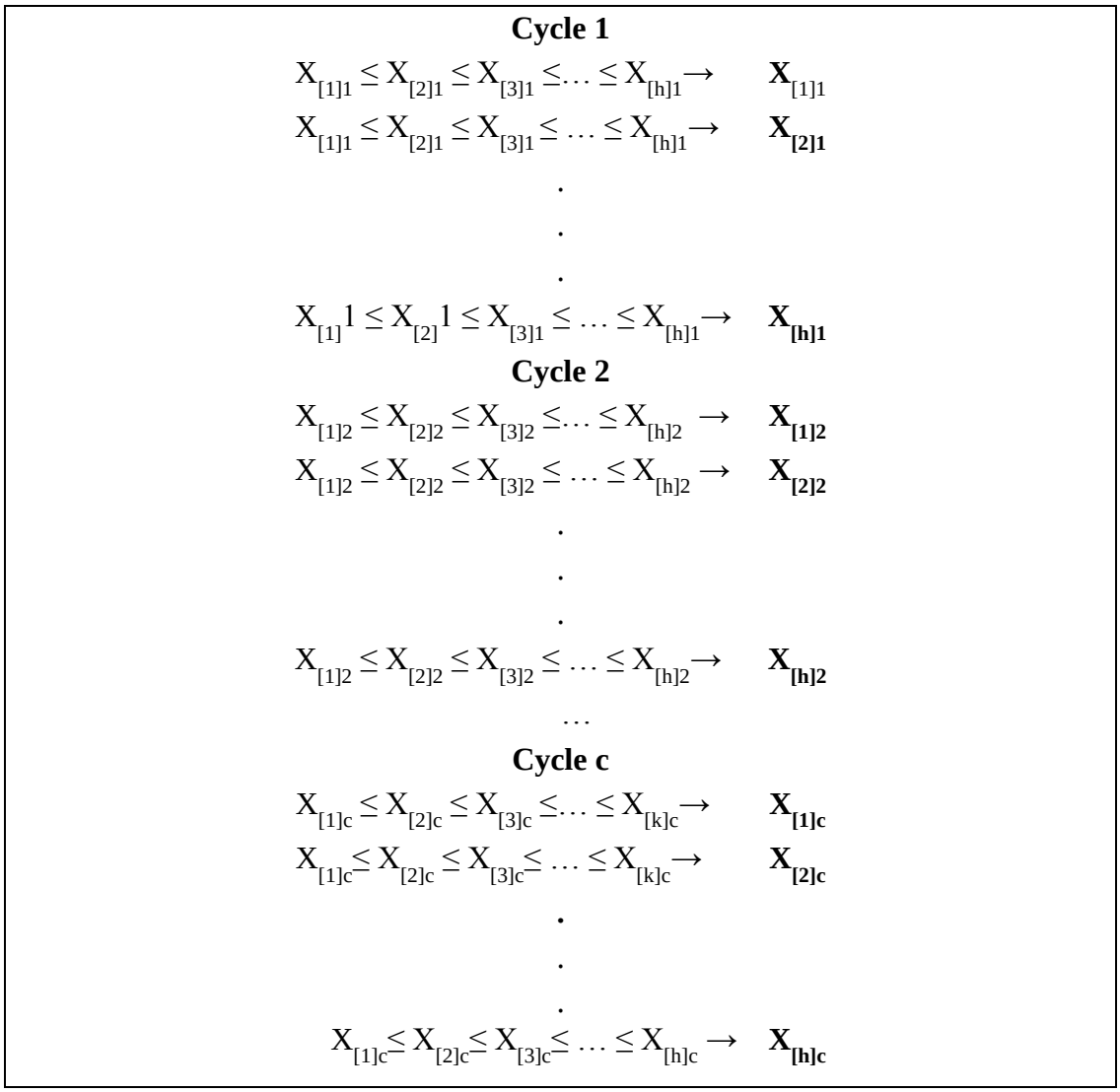


Figure 2.2 RSS procedure

The necessary mathematical background has been studied by Wolfe (2012). It has been showed by him that the sample mean of the RSS is an unbiased estimator of the population mean and the variance of ranked set sample mean is smaller than variance of simple random sample mean by equally size. Dell & Clutter (1972) stated that the RSS is the most useful technique sampling to drawn sampling when measurement of the object sampling is difficult but it can be sampled and ranked by a judgment or concomitant variable. Dell & Clutter (1972) also showed that efficiency of the mean estimator in RSS is better than the mean estimator in SRS by Using relative precession (RP).

The population arithmetic mean estimator for RSS which is denoted by μ_{RSS} can be seen in Equation 2.6. The variance of this estimator is given in Equation 2.7.

$$\hat{\mu}_{RSS} = \frac{1}{h}\sum_{i=1}^{h}\frac{1}{c}\sum_{j=1}^{c}X_{[i]j} = \frac{1}{h}\sum_{i=1}^{h}\bar{X}_{[i]} \tag{2.6}$$

Proof, $E(\hat{\mu}_{RSS}) = \mu$

$$E(\hat{\mu}_{RSS}) = E\left(\frac{1}{h}\sum_{i=1}^{h}\frac{1}{c}\sum_{j=1}^{c}X_{[i]j}\right) = \frac{1}{h}\sum_{i=1}^{h}\frac{1}{c}\sum_{j=1}^{c}E\left(X_{[i]j}\right) = \frac{1}{h}\sum_{j=1}^{h}E\left(X_{[i]}\right)$$

$$= \frac{1}{h}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}x\frac{h!}{(i-1)!(h-i)!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)$$

$$= \frac{1}{h}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}hx\frac{(h-1)!}{(i-1)!((h-1)-(i-1))!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)$$

$$= \frac{1}{h}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}hx\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)$$

$$= \int_{-\infty}^{\infty}xf(x)\left(\sum_{i=1}^{h}\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}\right)dx$$

since it is known from binomial distribution that

$$\sum_{i=1}^{h}\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i} = 1$$

Hence

$$E(\hat{\mu}_{RSS}) = \int_{-\infty}^{\infty}xf(x)dx = \mu$$

$$Var(\hat{\mu}_{RSS}) = Var(\bar{X}) - \frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2 \tag{2.7}$$

$$Var(\hat{\mu}_{RSS}) = Var\left(\frac{1}{h}\sum_{i=1}^{h}X_{[i]}\right) = \frac{1}{h^2}\sum_{i=1}^{h}Var\left(X_{[i]}\right) = \frac{1}{h^2}\sum_{i=1}^{h}E\left[\left(X_{[i]}-\mu_{[i]}\right)^2\right] \tag{2.8}$$

$$E\left[\left(X_{[i]}-\mu\right)^2\right] = E\left[\left(\left(X_{[i]}-\mu_{[i]}\right)+\left(\mu_{[i]}-\mu\right)\right)^2\right]$$

since the cross product terms are zero, the equation becomes:

$$E\left[\left(X_{[i]}-\mu\right)^2\right]=E\left[\left(X_{[i]}-\mu_{[i]}\right)^2\right]+\left(\mu_{[i]}-\mu\right)^2$$

so the equation can be constructed as:

$$E\left[\left(X_{[i]}-\mu_{[i]}\right)^2\right]=E\left[\left(X_{[i]}-\mu\right)^2\right]-\left(\mu_{[i]}-\mu\right)^2 \tag{2.9}$$

combining equation 2.8 and 2.9 yields the expression:

$$Var\left(\hat{\mu}_{RSS}\right)=\frac{1}{h^2}\sum_{i=1}^{h}E\left[\left(X_{[i]}-\mu\right)^2\right]-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

$$=\frac{1}{h^2}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}(x-\mu)^2\frac{h!}{(i-1)!(h-i)!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

$$=\frac{1}{h^2}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}h(x-\mu)^2\frac{(h-1)!}{(i-1)!((h-1)-(i-1))!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

$$=\frac{1}{h^2}\sum_{i=1}^{h}\left(\int_{-\infty}^{\infty}h(x-\mu)^2\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}f(x)dx\right)-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

$$=\frac{1}{h}\int_{-\infty}^{\infty}(x-\mu)^2f(x)\left(\sum_{i=1}^{h}\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}\right)dx-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

since it is known from binomial distribution that

$$\sum_{i=1}^{h}\binom{h-1}{i-1}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{h-i}=1$$

hence

$$Var(\bar{\mu}_{RSS})=\frac{1}{h}\int_{-\infty}^{\infty}(x-\mu)^2f(x)dx-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

and

$$\int_{-\infty}^{\infty}(x-\mu)^2f(x)dx=\sigma^2$$

so that,

$$Var(\bar{\mu}_{RSS})=\frac{\sigma^2}{h}-\frac{1}{h^2}\sum_{i=1}^{h}\left(\mu_{[i]}-\mu\right)^2$$

$$Var(\bar{\mu}_{RSS}) = Var(\bar{X}) - \frac{1}{h^2}\sum_{i=1}^{h}(\mu_{[i]} - \mu)^2$$

The effectiveness of RSS is related to how perfect the correlation between concomitant variable and the response variable. The best way to measure it is by developing the statistical correlation model to catch the uncertainty of the ranking process. Dell and Clutter (1972) have proposed one model to measure how perfect data are correlated. Their correlation model is using random error term (ε) by generating normal distribution with parameter Z(0,σ_ε^2) and it can be calculated by Equation 2.10 as follow:

$$\rho = \frac{1}{\sqrt{1+\frac{\sigma_\varepsilon^2}{\sigma_y^2}}} \tag{2.10}$$

Where ρ is the correlation coefficient and $\sigma_y^2$is the variance of the data. Therefore the correlation between response variable Y and concomitant variable X can be written as Equation 2.11.

$$X = Y + \varepsilon \tag{2.11}$$

The process to obtain the Equation 2.10 can be explained as follow.

$$\rho = \frac{Cov(X,Y)}{\sqrt{\sigma_y^2}\sqrt{\sigma_x^2}} \tag{2.12}$$

$$\begin{aligned} Cov(X,Y) &= E(YX) - E(Y)E(X) \\ &= E(Y(Y+\varepsilon)) - E(Y)E(Y+\varepsilon) \\ &= E(Y^2) + E(\varepsilon) - [E(Y)]^2 + E(Y).E(\varepsilon) \\ &= E(Y^2) + 0 - [E(Y)]^2 + E(Y).0 \\ &= E(Y^2) - [E(Y)]^2 = \sigma_y^2 \end{aligned} \tag{2.13}$$

Since response variable and random error term (ε) are independence and

$$Cov(Y,\varepsilon)=0 \text{ so}$$

$$\sigma^2(X)=\sigma^2(Y+\varepsilon)$$

$$=\sigma^2(Y)+\sigma^2(\varepsilon)+2Cov(Y,\varepsilon)$$

$$=\sigma^2(Y)+\sigma^2(\varepsilon)+0$$

$$=\sigma^2(Y)+\sigma^2(\varepsilon)$$

Substitute Equation 2.13 to Equation 2.12 yield the expression:

$$\rho=\frac{\sigma_y^2}{\sqrt{\sigma_y^2}\sqrt{\sigma_x^2}}=\frac{\sigma_y^2}{\sqrt{\sigma_y^2}\sqrt{\sigma_y^2+\sigma_\varepsilon^2}}=\frac{\sigma_y^2}{\sqrt{\sigma_y^4+\sigma_y^2\sigma_\varepsilon^2}}=\frac{\sigma_y^2}{\sigma_y^2\sqrt{1+\frac{\sigma_\varepsilon^2}{\sigma_y^2}}}=\frac{1}{\sqrt{1+\frac{\sigma_\varepsilon^2}{\sigma_y^2}}}$$

In order to do comparison, the new correlation model is developed by the regression concept. At this model of correlation, random error term (ε) by generating normal distribution with parameter $Z(0,\sigma_\varepsilon^2)$ is used so the expression between X and Y can be written as Equation 2.14.

$$X=Y-\varepsilon \tag{2.14}$$

The correlation coefficient and variance of data can be calculated by Equation 2.15 as follow:

$$\rho=\sqrt{1-\left(\frac{n-2}{n}\right)\frac{\sigma_\varepsilon^2}{\sigma_y^2}} \tag{2.15}$$

In regression, there are some definition like sum square error (SSE) and sum square total (SST). The correlation coefficient ρ is calculated by using SSR and SST and it is given in Equation 2.16.

$$\rho=\sqrt{1-\frac{SSE}{SST}} \tag{2.16}$$

And the definition SSE and SST can be described in Equation 2.17 and 2.18 respectively.

$$SST = n\sigma_y^2 \tag{2.17}$$

$$SSE = (n-2)\sigma_\varepsilon^2 \tag{2.18}$$

Substituting Equation 2.17 and 2.18 to Equation 2.16 can yield Equation 2.15. The process to derive Equation 2.15 is explained as below:

$$\rho = \sqrt{1 - \frac{SSE}{SST}}$$

$$= \sqrt{1 - \frac{(n-2)\sigma_\varepsilon^2}{n\sigma_y^2}}$$

$$= \sqrt{1 - \left(\frac{n-2}{n}\right)\frac{\sigma_\varepsilon^2}{\sigma_y^2}}$$

Considering that the value of $\left(\frac{n-2}{n}\right)$ is very small so it can be neglected.

CHAPTER THREE
MEAN ESTIMATOR FOR THE POPULATION HARMONIC MEAN AND GEOMETRIC MEAN IN RSS

Mean is technical term that is used to define central value. Generally, arithmetic mean is a central value that is used for calculating measure of location. Yet, using the arithmetic mean is not always an appropriate way to calculate a central value for some situation of data. There are the other means that can be used to calculate a central tendency for the certain situation of data such as geometric mean (GM) and harmonic mean (HM). In this chapter, we introduce the definitions of geometric and harmonic mean. Then we propose the new estimators for these population mean in RSS.

3.1 Geometric Mean

Misra and Gupta (2010) stated that there is condition where GM is more proper used than Arithmetic mean (AM), for instance data with positively skewed distribution or right skewed distribution. GM is specifically can be used as a central value in averaging ratio, percentages and rates of change in one period over the other. According to mathematical definition, GM is the n^{th} root of the product of n numbers. The geometric mean of N units of the population data can be calculated by Equation 3.1.

$$G = \sqrt[N]{X_1.X_2 \dots X_N} \tag{3.1}$$

Spizman (2008) stated that if the purpose of the growth rate is to account for compounding, the geometric mean of historical data should be employed. The common applications or growth rates are projecting the future losses of wages, the future costs of life care plan, and determining discount rates for present value calculations. Regardless which data application is used for estimating the growth rates, the value of AM is always greater than values of geometric mean.

3.2 Harmonic Mean

Besides the arithmetic mean and geometric mean, harmonic mean can be employed also to compute central value. The definition of HM is a central value used mainly in certain condition for instance when data consists of a set of rates, such as prices ($/kilo), speeds (mph), or productivity (output/man hour). The harmonic mean of N units of the population is calculated by Equation 3.2.

$$H = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\cdots+\frac{1}{X_N}} \tag{3.2}$$

Harmonic mean is sensitive to the outliers when they have much smaller values than the rest of the data , and largely insensitive to outliers when they have much larger value than the other data. The criterion for choosing between arithmetic mean and harmonic mean in order to calculate mean for rate is proposed by Ferger (1931) and Francis (2004). They stated that harmonic mean is proper if the rates are being averaged over constant numerator units and arithmetic mean can be used when the rates are being averaged over constant denominator units.

Burks (1985) showed that the HM of a random variable is always less than its GM, and GM is always less than its AM , or its relationship can be expressed as

$$\text{HM} \leq \text{GM} \leq \text{AM}$$

Interesting relationship also shown by Rossman (1990) that harmonic, geometric, and arithmetic mean in lognormal distributions is given in by Equation 3.3.

$$H = \frac{G^2}{\mu} \tag{3.3}$$

3.3 Proposed Estimator HM and GM in Ranked Set Sampling

In this subchapter, the estimators of HM and GM in ranked set sampling design are proposed in Equation 3.4 for harmonic mean and Equation 3.5 for geometric mean. It is also proved that these estimators are unbiased estimators.

3.3.1 Proposed Estimator of HM

$\overline{X}_{HRSS}$: The estimator of harmonic mean in RSS design

c : cycle size

h : set size

$$\overline{X}_{HRSS} = \frac{c.h}{\sum_{i=1}^{h}\sum_{j=1}^{c}\frac{1}{X_{[i]j}}} \tag{3.4}$$

3.3.2 Proposed Estimator of GM

$\overline{X}_{GRSS}$: The estimator of geometric mean in RSS design

c : cycle size

h : set size

$$\overline{X}_{GRSS} = \sqrt[c.h]{\prod_{i=1}^{c}\prod_{j=1}^{h} X_{[j]i}} \tag{3.5}$$

CHAPTER FOUR

SIMULATION STUDY

In this chapter, we deal with simulation study. This study consists of two parts. These are simulation process and simulation result. In simulation process, process of generating data in which two different correlation coefficient method is described and process of data simulation is explained. The Correlation coefficients are used namely the correlation coefficient from the imperfect judgment ranking model of Dell & Clutter (1972) and Adam & Levine (1972), and the correlation coefficient model of Dell & Clutter (1972) with modification based on regression concept. In simulation result part, results are shown by tables and graphics.

4.1 Simulation Process

We tried to show that estimator of Harmonic Mean (HM) and Geometric Mean (GM) in ranked set sampling is more efficient than estimator of HM and GM in simple random sampling by using a simulation study. The biases and relative efficiencies of these estimators are used. In order to check the efficiencies, data sets are generated from some distributions. In this study in order to obtain comparison the efficiencies based on the skewness of distribution, as can be seen in Figure 4.1, normal, lognormal, gamma, and exponential distribution are selected.

In this study, because of data that would be generated is bivariate data. And then correlation coefficient was involved to give measurement for the relationship between bivariate data (x, y). Here, two correlation coefficient models are used. First, the Dell and Clutter's model with modification by regression concept is used. Second, the imperfect judgment ranking model of Dell & Clutter (1972) and Adam & Levine (1972) is used.

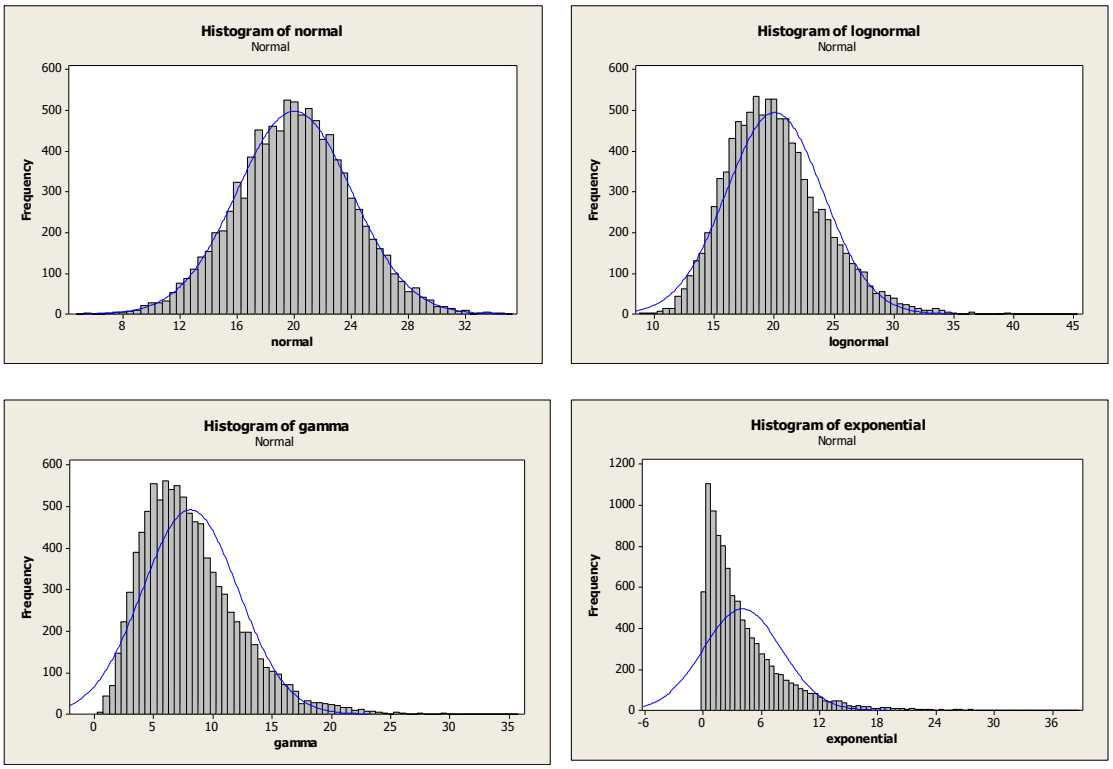


Figure 4.1 Histogram of generating data from different distributions.

The process of first method is explained step by step as this below:

1. Generate Y_i data from the distributions, at this simulation the distributions are normal(20,16), lognormal(2.9761,0.1981), gamma(4,2), and exponential(4). And the variance of this distribution is same each other as $\sigma^2 = 16$ for minimizing error that can be caused by variance in this study.

2. Generate random error term (ε) by generating normal distribution with parameter $Z(0,\sigma_\varepsilon^2)$, and σ_ε^2 was calculated by using Equation 4.1.

$$\rho = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}} \tag{4.1}$$

Where ρ = 0.5, 0.75, 0.9, 1. In order to measure perfect rate of relationship of data.

After the calculation σ_ε^2 is obtained as follows:

ρ=0.5 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 12$

ρ=0.75 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 7$

ρ=0.9 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 3.04$

ρ=1 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 0$

3. X_i is calculated by Equation 4.2.

$$X_i = Y_i - \varepsilon_i \tag{4.2}$$

The process of second method is explained step by step as this below:

1. Generate Y_i data from the distributions, at this simulation the distributions are normal(20,16), lognormal(2.9761,0.1981), gamma(4,2), and exponential(4). And the variance of this distribution is same each other as $\sigma^2 = 16$ for minimizing error that can be caused by variance in this study.

2. Generate random error term (ε) by generating normal distribution with parameter $Z(0,\sigma_\varepsilon^2)$, and σ_ε^2 was calculated by using equation 4.3.

$$\rho = \frac{1}{\sqrt{1+\frac{\sigma_\varepsilon^2}{\sigma_y^2}}} \tag{4.3}$$

Where ρ = 0.5, 0.75, 0.9, 1. In order to measure perfect rate of relationship of data.

After the calculation σ_ε^2 is obtained as follows:

ρ=0.5 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 48$

ρ=0.75 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 12.44$

ρ=0.9 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 3.753$

ρ=1 and $\sigma_y^2 = 16$ give the value $\sigma_\varepsilon^2 = 0$

3. X_i was calculated by equation 4.4.

$$X_i = Y_i + \varepsilon_i \tag{4.4}$$

Simulation process is done by using Matlab and Microsoft Excel. Generating data and calculation of harmonic mean, geometric mean, arithmetic mean and also standard deviation of these mean for both of ranked set sampling (RSS) and simple random sampling (SRS) is computed by using Matlab (The code of program is giving in Appendix). Then calculation of mean square error, bias and relative efficiency are computed by Microsoft excel.

For detail information, the process of simulation is explained as follows:

1. Generate 10.000 bivariate data (from normal, lognormal, gamma, exponential distribution and correlation coefficient ρ=0.5, 0.75, 0.9, 1).

2. After data is obtained, data is simulated 10.000 times for simple random sampling (SRS) method and rank set sampling (RSS) for every combination, N=c. h, where Cycle Size (c)= 2, 4, 10 and set Size (h)= 3, 4, 5.

3. After sampling mean data AM, HM, GM are obtained for every combination cycle size and set size, standard deviation of these mean are calculated one by one.

4. In order to calculate Mean Square Error (MSE), bias are computed and MSE is computed by equation 4.5.

$$\text{MSE} = \text{bias}^2 + \text{variance} \tag{4.5}$$

5. After MSE of each combination of cycle size and set size are obtained, MSE is used for calculating Relative Efficiency (RE). It is calculated by equation 4.6:

$$RE_{HM} = \frac{MSE_{SRS_{HM}}}{MSE_{RSS_{HM}}} \ , \ RE_{AM} = \frac{MSE_{SRS_{AM}}}{MSE_{RSS_{AM}}} \ , \ RE_{GM} = \frac{MSE_{SRS_{GM}}}{MSE_{RSS_{GM}}} \tag{4.6}$$

4.2 Simulation Result

In this section, results of simulation study are shown and described. First, the result of simulation process by using the correlation coefficient model of Dell & Clutter (1972) with modification based on regression concept is described. And then the result of simulation process by using the correlation coefficient from the imperfect judgment ranking model of Dell & Clutter (1972), and Adam & Levine (1972) is described later.

Results of simulation process are printed in table of relative efficiency, graphic of comparison of relative efficiency according to correlation coefficient, and table of biases.

4.2.1 Simulation Result of Simulation Process by Using The Correlation Coefficient Model of Dell & Clutter with Modification Based on Regression

In this section result from simulation process by using the correlation coefficient model of Dell & Clutter (1972) with modification based on regression concept is explained. The results are given by Table 4.1-4.4 as the Relative Efficiencies (RE) of AM, HM, and GM for normal, lognormal, gamma, and exponential distributions respectively. In these tables, the first column is the cycle size (c) of ranked set sampling with values 2, 6 and 10 respectively. The second column is information about set size (h) of ranked set sampling with values 3, 4 and 5. The third and the seventh are coefficient of correlation (ρ)with values 0.5, 0.75, 0.8 and 1. The values of relative efficiencies are shown in columns fourth and eight for arithmetic mean (AM), columns fifth and ninth for harmonic mean (HM), and columns sixth and tenth for geometric mean (GM). From the given tables, the comparison graphics are constructed in order to understand the result of relative efficiency easily. The comparison graphics is given by Figure 4.2-4.5 as the graphics of comparison of relative efficiency for normal, lognormal, gamma and exponential distributions respectively. The x-axis in graphics is the coefficient of correlation and the y-axis is the value of relative efficiencies of its coefficient of correlation. There are 9 graphics in these figures where the first row of every figure is the graphics for arithmetic

mean (AM) and the cycle size are 2, 4, and 6. The second row represents the graphics for harmonic mean (HM) and the cycle sizes are 2, 4, and 6. And the third row indicates graphics for geometric mean (GM) and the cycle size are 2, 4, 6. When we are looking at the graphics, there are 3 lines which are assigned by dot, square and diamond. The lines with dot are for set size 3, the lines with square are for set size 4, and the lines with diamond are for set size 5. Table biases are given also by Table 4.5 and Table 4.6 to provide the information for bias value of estimators.

According to Figure 4.2 and Figure 4.4 which are the graphics of simulation from normal and lognormal distributions, the results of simulation are like what we were expecting. It can be seen that every value of relative efficiency is more than 1 and it show that the estimators based on RSS are more efficient than the estimator based on SRS. From the Figure 4.2 and Figure 4.4 also can be indicated that when the coefficient of correlations is increasing then the value of relative efficiency also increases for AM, GM, and HM. Then the greater the value of set size yields the greater value of relative efficiency for AM, GM, and HM. It means the coefficient of correlation and set size affect the relative efficiency of simulation result for estimator of AM, HM, and GM.

The graphics of simulation result from gamma and exponential distributions are shown in Figure 4.4 and Figure 4.5. Generally, it can be stated that every value of relative efficiency is more than 1. Because of it, the estimators by using RSS design give more efficiency than estimators by using SRS design. For AM based on the Figure 4.4 and Figure 4.5 the simulation results from gamma and exponential distributions are as good as simulation result from normal and lognormal distribution. For GM, the coefficient of correlation and set size slightly affect the relative efficiency because when the values of the coefficient of correlation and set size are increasing then the value of relative efficiency increases but it is not as significant as the result from normal and lognormal distributions. For HM, if the data comes from exponential distribution it can be indicated that the values of coefficient of correlation from 0.5 to 0.9 do not affect the relative efficiency but the relative efficiency increases slightly when the coefficient of correlation is 1. For HM in exponential distribution, it can also be seen that set size does not affect the value of

relative efficiency because if the set size is increasing then relative efficiency does not always increase. For HM in simulation from gamma distribution, it can be stated that set size and coefficients of correlation from 0.5 to 0.9 slightly affect the value of the relative efficiency and the relative efficiency is increasing significantly when the coefficient of correlation is equal to 1. It is because when set size and coefficients of correlation from 0.5 to 0.9 are increasing then the relative efficiency also increases although the increments are not as significant as when the coefficient of correlation is 1.

Based on Table 4.5 and Table 4.6 which are tables of biases value, the values on these tables are very close to 0. It indicates that the estimators for AM, HM, and GM in RSS design are unbiased estimator.

Table 4.1 Relative Efficiencies (RE) of AM, HM, and GM for normal distribution.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.334	1.233	1.300	0.9	1.637	1.482	1.588
2	4	0.5	1.522	1.430	1.493	0.9	1.934	1.682	1.850
2	5	0.5	1.556	1.442	1.525	0.9	2.078	1.822	1.998
6	3	0.5	1.356	1.290	1.338	0.9	1.702	1.523	1.649
6	4	0.5	1.472	1.376	1.444	0.9	1.901	1.695	1.840
6	5	0.5	**1.621**	**1.483**	**1.583**	0.9	2.091	1.812	2.014
10	3	0.5	1.327	1.288	1.325	0.9	1.608	1.504	1.591
10	4	0.5	1.434	1.373	1.422	0.9	1.880	1.678	1.822
10	5	0.5	1.567	1.458	1.538	0.9	**2.138**	**1.859**	**2.063**
2	3	0.75	1.453	1.371	1.425	1	1.869	1.677	1.806
2	4	0.75	1.626	1.518	1.598	1	2.291	1.981	2.199
2	5	0.75	**1.804**	**1.647**	**1.762**	1	2.782	**2.286**	**2.631**
6	3	0.75	1.568	1.430	1.528	1	1.976	1.694	1.890
6	4	0.75	1.680	1.533	1.641	1	2.399	1.989	2.275
6	5	0.75	1.722	1.566	1.684	1	2.649	2.150	2.510
10	3	0.75	1.429	1.370	1.423	1	1.835	1.661	1.799
10	4	0.75	1.618	1.496	1.584	1	2.317	1.962	2.215
10	5	0.75	1.746	1.582	1.700	1	**2.725**	2.211	2.575

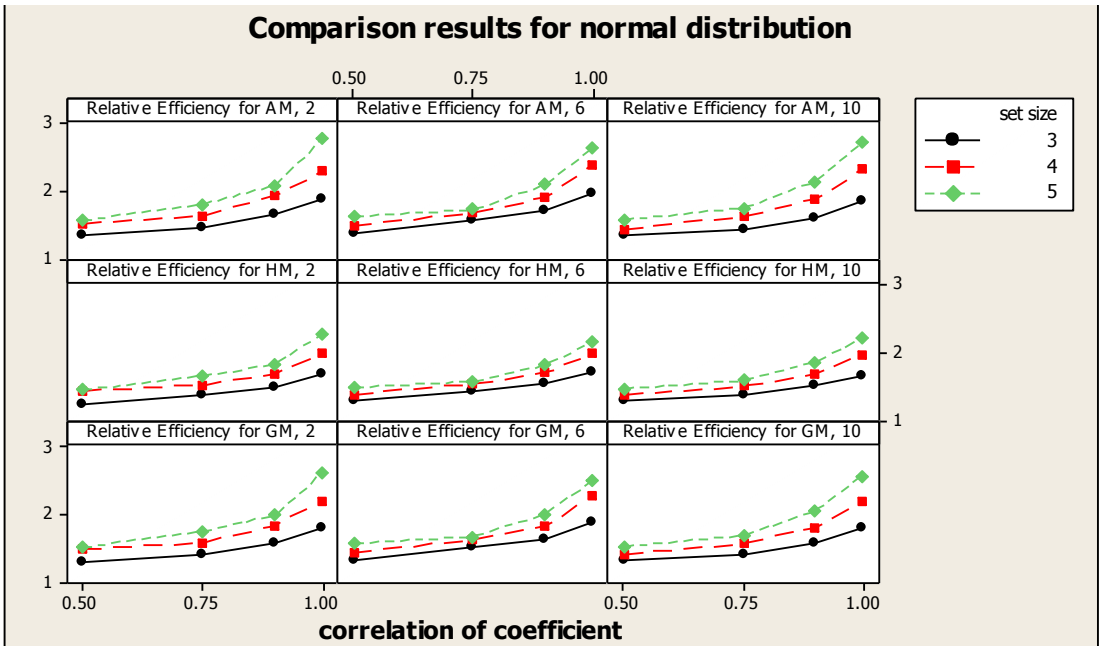


Figure 4.2 Graphics of comparison of relative efficiency for normal distribution.

Table 4.2 Relative Efficiencies (RE) of AM, HM, and GM for lognormal distribution.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.313	1.300	1.315	0.9	1.655	1.647	1.671
2	4	0.5	1.435	1.426	1.442	0.9	1.814	1.805	1.841
2	5	0.5	1.562	1.551	1.574	0.9	2.115	2.093	2.147
6	3	0.5	1.427	1.411	1.432	0.9	1.589	1.601	1.621
6	4	0.5	1.482	1.436	1.472	0.9	1.869	1.833	1.885
6	5	0.5	1.493	1.468	1.496	0.9	2.077	2.043	2.107
10	3	0.5	1.303	1.312	1.318	0.9	1.627	1.605	1.639
10	4	0.5	1.432	1.411	1.433	0.9	1.927	1.875	1.937
10	5	0.5	1.508	1.475	1.507	0.9	2.163	2.125	2.193
2	3	0.75	1.457	1.456	1.472	1	1.832	1.855	1.879
2	4	0.75	1.582	1.572	1.597	1	2.216	2.270	2.299
2	5	0.75	1.750	1.728	1.766	1	2.694	2.735	2.805
6	3	0.75	1.441	1.416	1.442	1	1.938	1.935	1.981
6	4	0.75	1.648	1.648	1.672	1	2.344	2.331	2.403
6	5	0.75	1.757	1.714	1.761	1	2.549	2.585	2.655
10	3	0.75	1.404	1.420	1.426	1	1.780	1.828	1.840
10	4	0.75	1.601	1.565	1.602	1	2.255	2.262	2.321
10	5	0.75	1.758	1.712	1.762	1	2.625	2.645	2.727

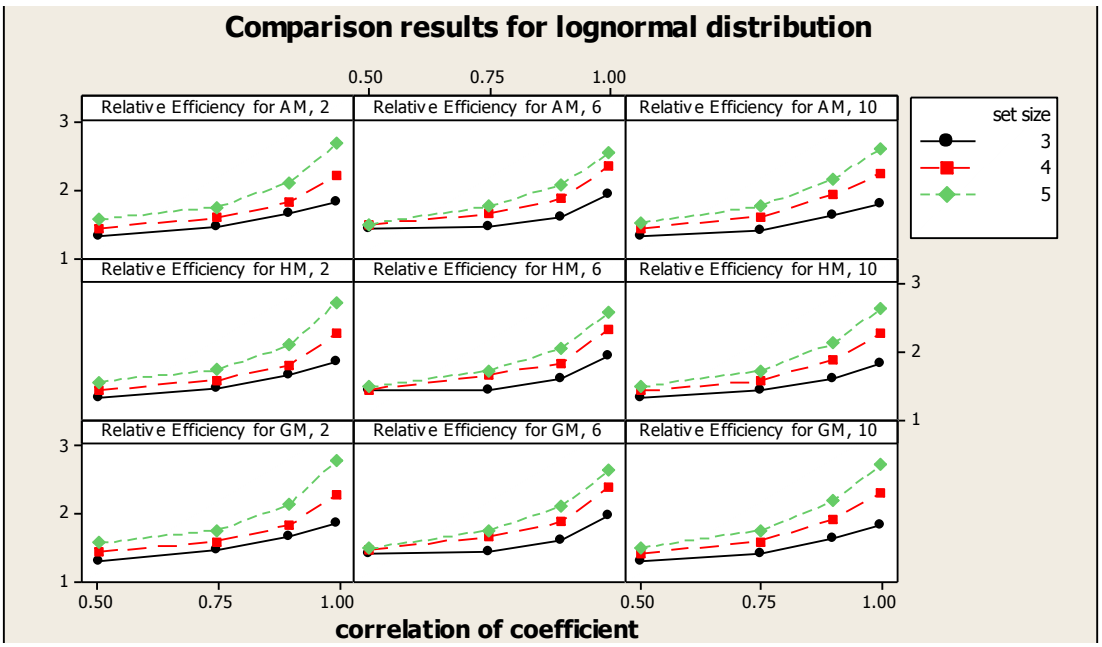


Figure 4.3 Graphics of comparison of relative efficiency for lognormal distribution.

Table 4.3 Relative Efficiencies (RE) of AM, HM, and GM for gamma distribution.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.345	1.273	1.346	0.9	1.557	1.439	1.599
2	4	0.5	1.433	1.323	1.441	0.9	1.761	1.548	1.813
2	5	0.5	1.512	1.316	**1.469**	0.9	2.015	**1.610**	**2.011**
6	3	0.5	1.354	1.221	1.344	0.9	1.621	1.401	1.648
6	4	0.5	1.399	1.265	1.390	0.9	1.847	1.477	1.846
6	5	0.5	1.506	**1.273**	1.464	0.9	2.038	1.508	1.993
10	3	0.5	1.291	1.180	1.295	0.9	1.655	1.350	1.661
10	4	0.5	1.360	1.200	1.346	0.9	1.772	1.430	1.799
10	5	0.5	**1.522**	1.237	1.456	0.9	**2.063**	1.461	2.001
2	3	0.75	1.436	1.352	1.458	1	1.821	1.683	1.929
2	4	0.75	1.543	1.401	1.569	1	2.105	1.882	2.286
2	5	0.75	1.696	**1.449**	1.684	1	2.524	**2.075**	**2.716**
6	3	0.75	1.425	1.299	1.443	1	1.818	1.588	1.931
6	4	0.75	1.551	1.330	1.522	1	2.148	1.757	2.273
6	5	0.75	1.695	1.348	1.645	1	2.508	1.888	2.666
10	3	0.75	1.405	1.238	1.404	1	1.747	1.470	1.818
10	4	0.75	1.523	1.284	1.503	1	2.019	1.684	2.169
10	5	0.75	**1.753**	1.363	**1.692**	1	**2.490**	1.895	2.660

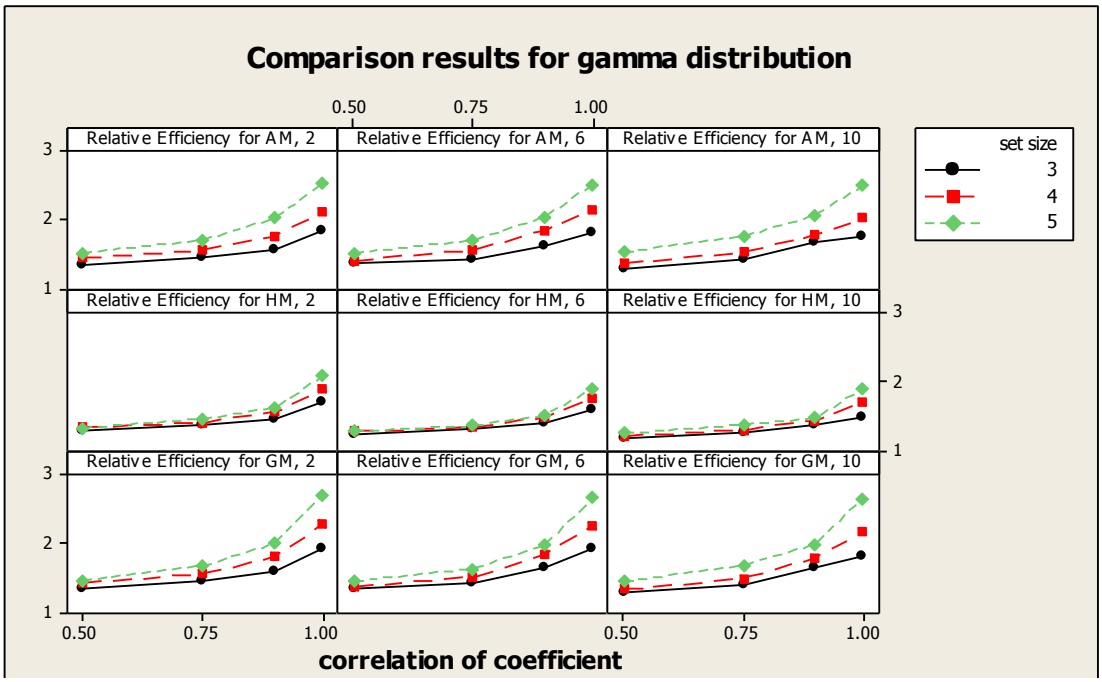


Figure 4.4 Graphics of comparison of relative efficiency for gamma distribution.

Table 4.4 Relative Efficiencies (RE) of AM, HM, and GM for exponential distribution.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.278	**1.157**	**1.298**	0.9	1.430	**1.276**	1.531
2	4	0.5	1.348	1.128	1.295	0.9	1.548	1.246	1.555
2	5	0.5	1.357	1.087	1.255	0.9	**1.827**	1.217	**1.679**
6	3	0.5	1.276	1.071	1.224	0.9	1.466	1.159	1.466
6	4	0.5	1.342	1.072	1.267	0.9	1.659	1.121	1.540
6	5	0.5	**1.496**	1.039	1.293	0.9	1.790	1.097	1.571
10	3	0.5	1.250	1.041	1.170	0.9	1.441	1.033	1.328
10	4	0.5	1.288	1.027	1.165	0.9	1.665	1.097	1.500
10	5	0.5	1.381	1.042	1.229	0.9	1.815	1.050	1.530
2	3	0.75	1.357	**1.208**	1.401	1	1.575	1.558	2.034
2	4	0.75	1.478	1.190	**1.438**	1	1.962	**1.731**	2.569
2	5	0.75	1.522	1.149	1.417	1	2.199	1.692	**2.818**
6	3	0.75	1.336	1.101	1.296	1	1.688	1.309	1.955
6	4	0.75	1.451	1.067	1.343	1	1.945	1.403	2.351
6	5	0.75	**1.623**	1.054	1.362	1	**2.222**	1.388	2.682
10	3	0.75	1.345	1.041	1.256	1	1.659	1.188	1.845
10	4	0.75	1.492	1.040	1.322	1	1.918	1.224	2.168
10	5	0.75	1.553	1.050	1.345	1	2.204	1.245	2.685

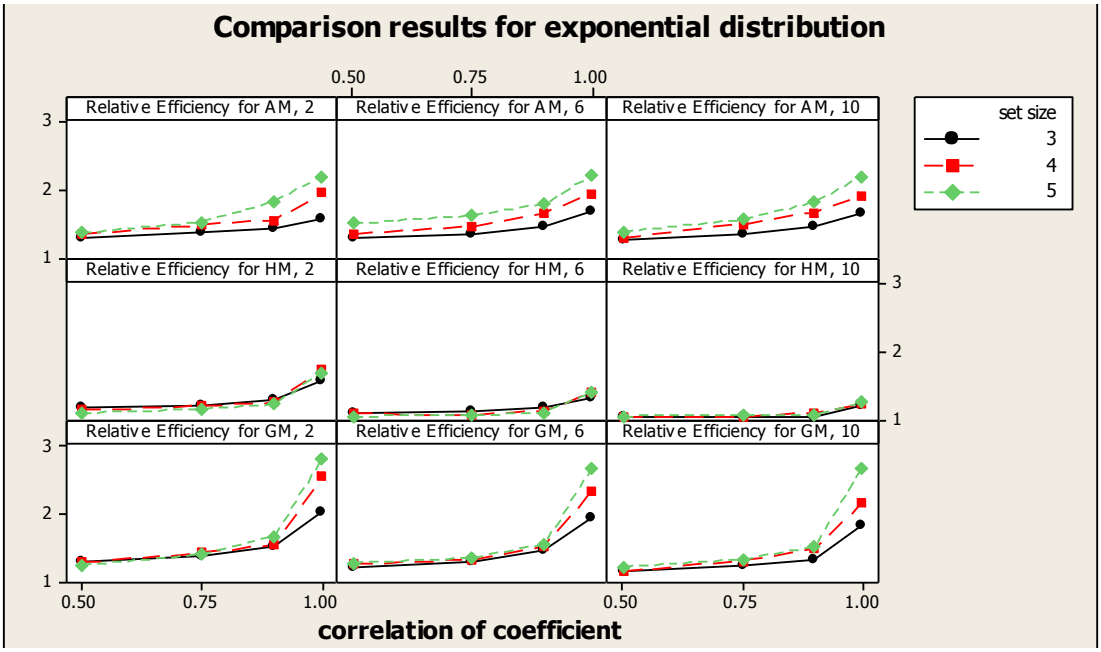


Figure 4.5 Graphics of comparison of relative efficiency for exponential distribution.

Table 4.5 Biases of AM,GM, and HM for normal and exponential distribution.

ρ	C	h	Normal (20,16)			Exp (4)		
			AM	GM	HM	AM	GM	HM
0.5	2	3	0.00126	0.000992	0.00059	1.69E-06	1.02E-05	0.000475
0.5	2	4	7.84E-06	1.60E-07	1.23E-05	0.00031	5.48E-05	1.44E-06
0.5	2	5	0.000259	0.000231	0.00019	1.09E-05	1.94E-05	2.56E-06
0.5	6	3	6.40E-07	9.00E-08	1.00E-08	4.90E-05	4.00E-06	6.89E-05
0.5	6	4	2.21E-05	9.00E-06	0	1.76E-05	4.76E-05	0
0.5	6	5	0.00019	0.00018	0.000151	1.16E-05	8.41E-06	6.24E-05
0.5	10	3	0.000135	0.000135	0.000144	9.60E-05	0.00011	7.29E-06
0.5	10	4	2.56E-06	4.41E-06	1.09E-05	3.60E-07	6.76E-06	4.90E-07
0.5	10	5	3.24E-06	2.25E-06	2.25E-06	1.96E-06	4.90E-07	0.000213
0.75	2	3	4.00E-08	3.60E-07	2.56E-06	0.00011	4.23E-05	1.96E-06
0.75	2	4	0.000182	0.000269	0.000357	0.000692	0.000119	0.000458
0.75	2	5	5.29E-06	5.29E-06	5.29E-06	1.00E-04	1.00E-06	6.40E-05
0.75	6	3	2.50E-05	2.03E-05	6.25E-06	4.00E-06	6.40E-07	1.22E-05
0.75	6	4	0.0001	9.22E-05	0.000102	4.00E-08	5.76E-06	9.02E-05
0.75	6	5	5.29E-06	1.52E-05	3.48E-05	9.60E-05	5.78E-05	0.000234
0.75	10	3	0.000161	0.000177	0.000204	2.21E-05	2.03E-05	2.92E-05
0.75	10	4	3.61E-06	2.56E-06	4.00E-06	1.68E-05	3.61E-06	7.92E-05
0.75	10	5	7.29E-06	2.40E-05	6.89E-05	7.29E-06	1.76E-05	0.000149
0.9	2	3	3.72E-05	9.00E-08	0.000119	9.60E-05	0.000104	0.001697
0.9	2	4	3.48E-05	0.000213	0.000767	0.000154	2.30E-05	0.000506
0.9	2	5	4.90E-07	5.48E-05	0.000279	8.41E-06	0.000139	0.000185
0.9	6	3	6.25E-06	0.000125	0.0006	0.000137	5.78E-05	6.56E-05
0.9	6	4	1.00E-06	3.72E-05	0.000182	3.72E-05	2.40E-05	0.000246
0.9	6	5	1.96E-06	6.89E-05	0.000388	1.76E-05	3.14E-05	0.000169
0.9	10	3	1.69E-06	3.36E-05	0.000306	2.30E-05	2.25E-06	2.56E-06
0.9	10	4	3.36E-05	0.000125	0.000462	1.60E-07	6.40E-07	8.46E-05
0.9	10	5	2.03E-05	8.46E-05	0.000296	1.96E-06	1.02E-05	0.000161
1	2	3	5.18E-05	4.76E-05	2.70E-05	0.001253	0.000562	0.0001
1	2	4	0.000123	0.000243	0.000396	0.00024	5.04E-05	2.21E-05
1	2	5	2.25E-06	7.84E-06	1.60E-05	3.36E-05	1.76E-05	0.0004
1	6	3	2.81E-05	2.21E-05	6.76E-06	1.30E-05	4.00E-08	4.90E-07
1	6	4	7.06E-05	7.23E-05	8.65E-05	7.84E-06	1.09E-05	6.40E-07
1	6	5	2.50E-07	1.09E-05	4.49E-05	2.12E-05	1.85E-05	3.25E-05
1	10	3	0.000115	0.000132	0.000164	3.60E-05	4.41E-06	3.24E-06
1	10	4	7.84E-06	6.25E-06	7.84E-06	6.76E-06	4.00E-08	4.00E-08
1	10	5	1.30E-05	3.03E-05	7.40E-05	1.44E-06	1.96E-06	5.04E-05

Table 4.6 Biases of AM,GM, and HM for gamma and lognormal distribution.

			Gamma (4,2)			LogNormal (2.9761, 0.1980)		
ρ	**C**	**h**	**AM**	**GM**	**HM**	**AM**	**GM**	**HM**
0.5	2	3	2.12E-05	0.12567	0.016333	6.40E-07	9.58E-03	0.002372
0.5	2	4	3.14E-05	6.64E-02	7.92E-03	0.000146	7.78E-03	2.49E-03
0.5	2	5	7.74E-05	0.042642	0.003795	2.03E-05	2.28E-03	4.54E-04
0.5	6	3	7.40E-05	1.86E-02	2.51E-03	3.14E-05	1.32E-03	4.37E-04
0.5	6	4	1.00E-08	6.96E-03	0.000534	1.04E-04	2.25E-04	5.76E-06
0.5	6	5	1E-06	0.005344	0.000548	7.84E-06	3.39E-04	1.08E-04
0.5	10	3	5.33E-05	0.00743	0.001096	1.23E-04	7.74E-05	1.21E-06
0.5	10	4	6.40E-07	3.83E-03	3.96E-04	1.60E-07	1.39E-04	2.92E-05
0.5	10	5	4.90E-07	2.49E-03	2.28E-04	6.76E-06	1.06E-04	1.3E-05
0.75	2	3	6.56E-05	1.11E-01	1.42E-02	0.000207	1.11E-02	3.55E-03
0.75	2	4	0.000106	0.061554	0.007621	4.9E-07	0.002704	0.00066
0.75	2	5	1.06E-04	3.19E-02	2.31E-03	1.44E-06	1.96E-03	4.24E-04
0.75	6	3	1.22E-05	1.68E-02	1.97E-03	6.25E-06	1.38E-03	3.92E-04
0.75	6	4	7.23E-05	8.84E-03	0.001109	1.12E-04	9.73E-04	4.20E-04
0.75	6	5	1.44E-05	5.17E-03	5.66E-04	1.06E-04	6.25E-04	0.000306
0.75	10	3	2.5E-05	0.006432	0.000853	1.44E-04	4.49E-05	7.29E-06
0.75	10	4	4.00E-08	3.49E-03	3.17E-04	1.68E-05	2.37E-04	9.02E-05
0.75	10	5	5.29E-06	2.38E-03	2.40E-04	4.00E-06	2.10E-04	6.4E-05
0.9	2	3	3.84E-05	1.26E-01	0.015154	2.22E-03	0.016129	0.007586
0.9	2	4	1.69E-04	0.055554	0.005256	0.001005	5.60E-03	0.002767
0.9	2	5	7.74E-05	3.58E-02	0.002411	1.96E-03	0.005506	0.003457
0.9	6	3	2.81E-05	0.018824	0.001927	0.00126	2.87E-03	1.97E-03
0.9	6	4	4.41E-06	8.35E-03	0.000502	1.59E-03	2.52E-03	0.002025
0.9	6	5	2.92E-05	5.73E-03	0.000475	1.62E-03	1.93E-03	0.001764
0.9	10	3	4.00E-08	6.96E-03	0.000548	1.20E-03	1.92E-03	1.53E-03
0.9	10	4	1.22E-05	0.00256	8.84E-05	1.78E-03	2.15E-03	1.96E-03
0.9	10	5	1.60E-05	1.89E-03	6.56E-05	1.69E-03	1.93E-03	0.001815
1	2	3	6.89E-03	5.33E-02	2.92E-04	3.97E-05	0.003832	0.000724
1	2	4	0.007259	0.020535	0.000292	4.23E-05	3.28E-03	1.00E-03
1	2	5	1.06E-02	5.40E-03	2.70E-03	8.10E-07	7.08E-04	0.000159
1	6	3	8.23E-03	3.56E-03	1.85E-03	2.70E-05	7.73E-04	2.66E-04
1	6	4	6.94E-03	8.47E-04	2.33E-03	8.65E-05	3.97E-05	1.96E-06
1	6	5	8.70E-03	1.22E-05	3.93E-03	2.89E-06	4.90E-05	1.76E-05
1	10	3	0.008575	0.000376	0.003147	1.12E-04	1.30E-05	1.22E-05
1	10	4	9.14E-03	1.37E-05	4.40E-03	1.23E-05	1.12E-04	4.62E-05
1	10	5	9.10E-03	9.60E-05	4.71E-03	4.00E-06	1.14E-04	3.97E-05

4.2.2 Simulation Result of Simulation Process by Using The Correlation Coefficient from The Imperfect Judgment Ranking Model of Dell & Clutter and Adam & Levine

In this section result from simulation process by using the correlation coefficient from the imperfect judgment ranking model of Dell & Clutter (1972) and Adam & Levine (1972) is explained. The results are given by Table 4.7-4.10 as the Relative Efficiencies (RE) of AM, HM, and GM for normal, lognormal, gamma, and exponential distributions respectively. In these tables, the first column is the cycle size (c) of ranked set sampling with values 2, 6 and 10 respectively. The second column is information about set size (h) of ranked set sampling with values 3, 4 and 5. The third and the seventh are coefficient of correlation (ρ)with values 0.5, 0.75, 0.8 and 1. The values of relative efficiencies are shown in columns fourth and eight for arithmetic mean (AM), columns fifth and ninth for harmonic mean (HM), and columns sixth and tenth for geometric mean (GM). From the given tables, the comparison graphics are constructed in order to understand the result of relative efficiency easily. The comparison graphics are given by Figure 4.6-4.9 as the graphics of comparison of relative efficiency for normal, lognormal, gamma and exponential distributions respectively. The x-axis in graphics is the coefficient of correlation and the y-axis is the value of relative efficiencies of its coefficient of correlation. There are 9 graphics in these figures where the first row of every figure is the graphics for arithmetic mean (AM) and the cycle size are 2, 4, and 6. The second row on the other hand represents the graphics for harmonic mean (HM) and the cycle sizes are 2, 4, and 6. And the third row indicates graphics for geometric mean (GM) and the cycle size are 2, 4, 6. When we are looking at the graphics, there are 3 line which are colored by black, red and green. The black lines are for set size 3, the red lines are for set size 4, and the green lines are for set size 5. Table biases are given also by Table 4.11 and Table 4.12 to provide the information for bias value of estimators.

Figure 4.6 and 4.7 are graphics of relative efficiency for simulation result from normal and lognormal distribution. According to these graphics, it can be seen that every value of relative efficiency is over than 1 and it indicates that the estimators in

RSS design is more efficiency than the estimators in SRS design. These graphics also indicate that when the value of correlation of coefficient is increasing then the value of relative efficiency also increases for AM, HM, and GM and it means the correlation of coefficient affects the value of relative efficiency. From these graphics, it can be state that set size affects the value of relative efficiency because when the set size is increasing then the value of relative efficiency increases.

The graphics of simulation result from gamma and exponential distributions are shown in Figure 4.8 and Figure 4.9. Generally, it can be concluded that the estimators of AM, HM, and GM in RSS are more efficient than the estimators of AM, HM, and GM in SRS because every value of relative efficiency is more than 1. For AM based on the Figure 4.8 and Figure 4.9 the simulation result from gamma and exponential distributions are as good as simulation result from normal and lognormal distribution. Therefore the set size and the correlation coefficient give effect for the value of relative efficiency for simulation of AM from gamma and exponential distributions. For GM, it can be stated that when the values of the coefficient of correlation and set size is increasing then the value of relative efficiency increases if the data comes from gamma distribution but if the data comes from exponential distribution the set size does not significantly affect the value of relative efficiency. For HM, if the data comes from exponential distribution it can be indicated that the set size and the values of coefficient of correlation from 0.5 to 0.9 do not affect the relative efficiency but the relative efficiency increases slightly when the coefficient of correlation is 1. For HM in simulation result from gamma distribution, it can be stated that the set size and the coefficients of correlation slightly affect the value of the relative efficiency although the effect of the coefficient of correlation and the set size are not as significant as the result from normal and lognormal distributions.

According to Table 4.11 and Table 4.12, it can be concluded that the estimators of AM, HM, and GM are unbiased estimator because every values on Table 4.11 and Table 4.12 are very close to 1.

Table 4.7 Relative Efficiencies (RE) of AM, HM, and GM for normal distribution from simulation process by using original Dell & Clutter model.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.218	1.123	1.141	0.9	1.690	1.464	1.549
2	4	0.5	1.278	1.131	1.149	0.9	1.917	1.655	1.774
2	5	0.5	1.290	1.155	1.167	0.9	2.049	1.823	1.981
6	3	0.5	1.402	1.157	1.171	0.9	1.994	1.536	1.636
6	4	0.5	1.455	1.129	1.133	0.9	2.341	1.634	1.769
6	5	0.5	1.624	**1.188**	**1.206**	0.9	2.807	**1.846**	**2.028**
10	3	0.5	1.529	1.118	1.139	0.9	2.123	1.502	1.585
10	4	0.5	1.673	1.120	1.132	0.9	2.690	1.664	1.783
10	5	0.5	**1.888**	1.168	1.189	0.9	**3.277**	1.817	1.997
2	3	0.75	1.430	1.287	1.329	1	2.039	1.705	1.851
2	4	0.75	1.606	1.369	1.427	1	2.383	1.910	2.150
2	5	0.75	1.753	1.472	1.552	1	2.708	2.243	2.566
6	3	0.75	1.696	1.359	1.408	1	2.254	1.645	1.799
6	4	0.75	1.857	1.363	1.422	1	3.012	1.993	2.263
6	5	0.75	2.127	1.468	**1.552**	1	3.657	**2.213**	**2.571**
10	3	0.75	1.791	1.278	1.327	1	2.569	1.703	1.868
10	4	0.75	2.125	1.384	1.431	1	3.396	2.008	2.293
10	5	0.75	**2.495**	**1.449**	1.538	1	**4.179**	2.126	2.476

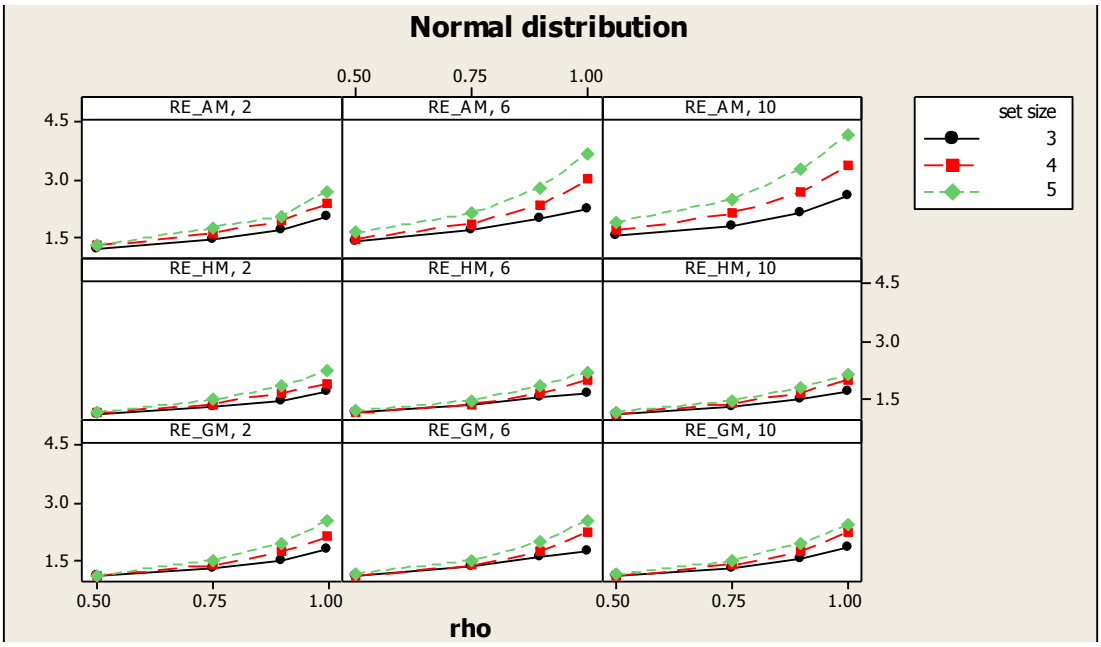


Figure 4.6 Graphics of comparison of relative efficiency for normal distribution from simulation process by using original Dell & Clutter model.

Table 4.8 Relative Efficiencies (RE) of AM, HM, and GM for lognormal distribution from simulation process by using original Dell & Clutter model.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.132	1.133	1.136	0.9	1.609	1.609	1.631
2	4	0.5	1.155	1.135	1.148	0.9	1.762	1.728	1.774
2	5	0.5	1.163	1.156	1.164	0.9	1.971	**1.966**	**2.009**
6	3	0.5	1.113	1.101	1.110	0.9	1.562	1.551	1.578
6	4	0.5	**1.203**	1.178	**1.194**	0.9	1.859	1.831	1.876
6	5	0.5	1.138	1.135	1.140	0.9	**1.976**	1.952	2.004
10	3	0.5	1.138	1.126	1.136	0.9	1.569	1.553	1.583
10	4	0.5	1.162	1.163	1.166	0.9	1.871	1.864	1.906
10	5	0.5	1.182	**1.167**	1.180	0.9	1.927	1.894	1.947
2	3	0.75	1.336	1.325	1.340	1	1.876	1.897	1.924
2	4	0.75	1.425	1.396	1.423	1	2.209	2.219	2.274
2	5	0.75	1.522	1.502	1.528	1	2.589	2.657	2.712
6	3	0.75	1.300	1.286	1.303	1	1.822	1.829	1.863
6	4	0.75	1.489	1.449	1.483	1	2.308	2.317	2.375
6	5	0.75	1.540	1.510	1.541	1	**2.610**	**2.645**	2.718
10	3	0.75	1.348	1.324	1.348	1	1.840	1.858	1.888
10	4	0.75	1.496	1.482	1.506	1	2.291	2.338	2.385
10	5	0.75	**1.556**	**1.531**	**1.561**	1	2.529	2.547	**2.620**

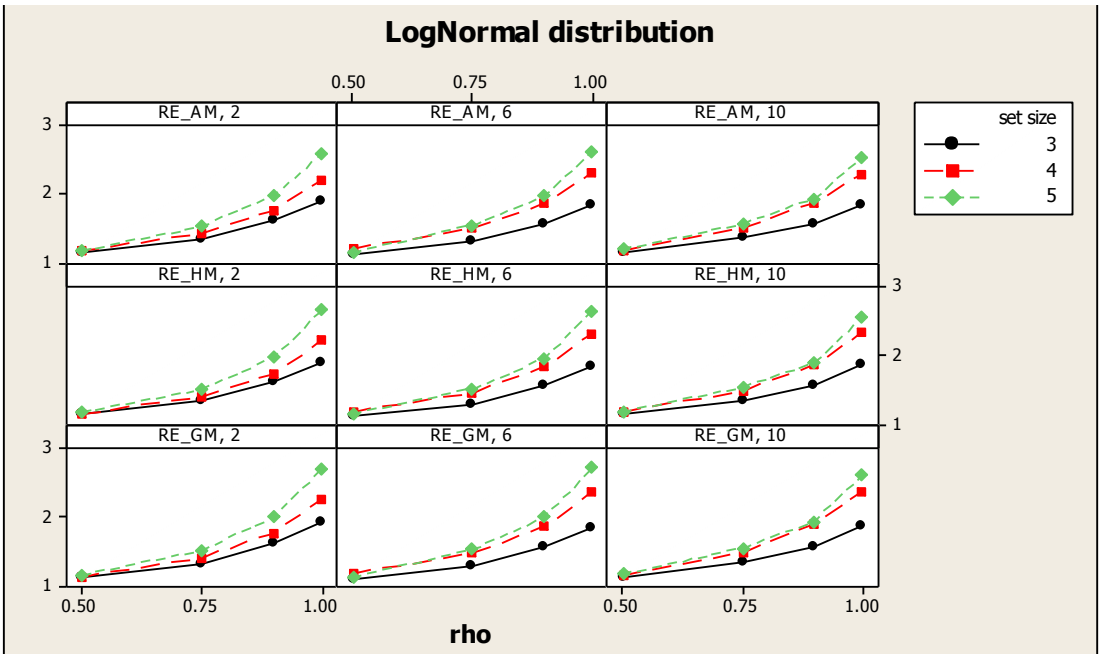


Figure 4.7 Graphics of comparison of relative efficiency for lognormal distribution from simulation process by using original Dell & Clutter model.

Table 4.9 Relative Efficiencies (RE) of AM, HM, and GM for gamma distribution from simulation process by using original Dell & Clutter model.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.120	1.111	1.126	0.9	1.560	1.452	1.590
2	4	0.5	1.146	1.127	1.155	0.9	1.733	1.542	1.770
2	5	0.5	1.160	1.101	1.143	0.9	1.940	**1.603**	1.925
6	3	0.5	1.106	1.058	1.091	0.9	1.549	1.330	1.543
6	4	0.5	1.174	1.096	1.157	0.9	1.802	1.460	1.784
6	5	0.5	**1.192**	**1.139**	**1.195**	0.9	**1.967**	1.544	**1.965**
10	3	0.5	1.137	1.045	1.107	0.9	1.579	1.309	1.546
10	4	0.5	1.127	1.074	1.113	0.9	1.747	1.421	1.725
10	5	0.5	1.141	1.078	1.129	0.9	1.945	1.550	1.938
2	3	0.75	1.334	1.247	1.332	1	1.802	1.674	1.901
2	4	0.75	1.456	1.268	1.421	1	2.140	1.868	2.292
2	5	0.75	1.559	1.336	1.522	1	2.535	**2.023**	2.667
6	3	0.75	1.403	1.258	1.390	1	1.816	1.530	1.880
6	4	0.75	1.472	1.252	1.433	1	2.238	1.763	2.358
6	5	0.75	1.579	**1.299**	1.500	1	2.573	1.932	2.745
10	3	0.75	1.355	1.188	1.332	1	1.846	1.493	1.879
10	4	0.75	1.459	1.209	1.404	1	2.161	1.681	2.251
10	5	0.75	**1.630**	1.290	**1.557**	1	**2.524**	1.917	**2.693**

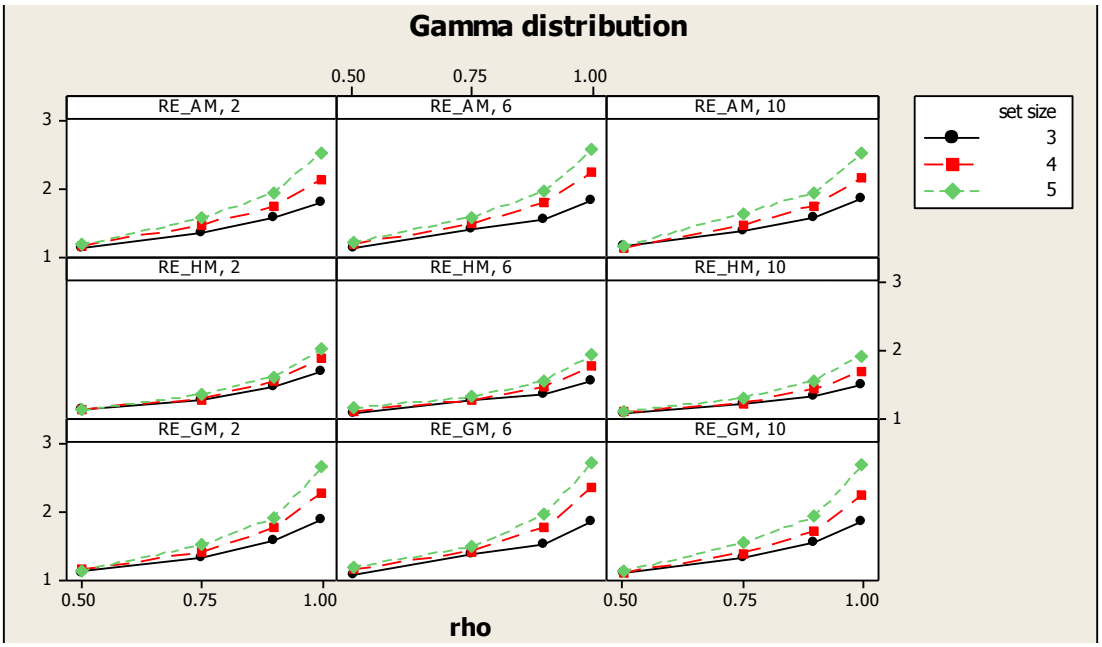


Figure 4.8 Graphics of comparison of relative efficiency for gamma distribution from simulation process by using original Dell & Clutter model.

Table 4.10 Relative Efficiencies (RE) of AM, HM, and GM for exponential distribution from simulation process by using original Dell & Clutter model.

cycle size	set size	ρ	AM	HM	GM	ρ	AM	HM	GM
2	3	0.5	1.088	1.039	1.097	0.9	1.383	1.265	1.507
2	4	0.5	1.181	**1.116**	**1.189**	0.9	1.659	**1.351**	1.759
2	5	0.5	1.196	1.046	1.141	0.9	1.764	1.265	**1.727**
6	3	0.5	1.174	1.026	1.124	0.9	1.488	1.144	1.477
6	4	0.5	1.168	1.087	1.148	0.9	1.623	1.165	1.559
6	5	0.5	1.182	1.046	1.142	0.9	**1.796**	1.115	1.634
10	3	0.5	1.123	1.025	1.073	0.9	1.458	1.108	1.430
10	4	0.5	1.151	1.007	1.095	0.9	1.619	1.056	1.484
10	5	0.5	**1.189**	1.029	1.141	0.9	1.773	1.086	1.631
2	3	0.75	1.232	1.135	1.275	1	1.575	1.558	2.034
2	4	0.75	1.394	**1.196**	**1.388**	1	1.962	1.731	2.569
2	5	0.75	1.465	1.144	1.385	1	2.199	**1.692**	**2.818**
6	3	0.75	1.315	1.072	1.272	1	1.688	1.309	1.955
6	4	0.75	1.380	1.112	1.310	1	1.945	1.403	2.351
6	5	0.75	**1.484**	1.067	1.345	1	**2.222**	1.388	2.682
10	3	0.75	1.288	1.057	1.222	1	1.659	1.188	1.845
10	4	0.75	1.366	1.020	1.244	1	1.918	1.224	2.168
10	5	0.75	1.441	1.045	1.313	1	2.204	1.245	2.685

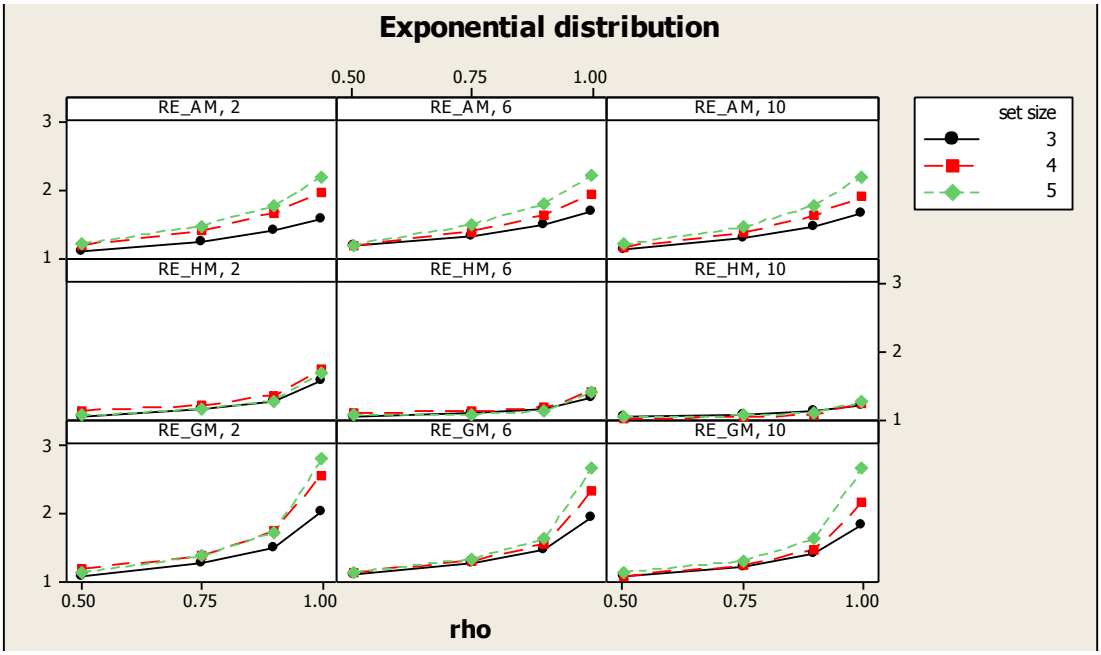


Figure 4.9 Graphics of comparison of relative efficiency for exponential distribution from simulation process by using original Dell & Clutter model.

Table 4.11. Biases of AM,GM, and HM for normal and exponential distribution from simulation process by using original Dell & Clutter model.

ρ	c	h	Normal (20,16)			Exp (4)		
			AM	GM	HM	AM	GM	HM
0.5	2	3	0.000686	0.029653	0.007726	1.16E-03	1.17E+00	0.08673
0.5	2	4	1.96E-06	1.23E-02	2.08E-03	0.000207	7.96E-01	3.66E-02
0.5	2	5	7.29E-06	0.008538	0.001354	3.36E-05	6.42E-01	2.79E-02
0.5	6	3	6.56E-05	2.30E-03	2.34E-04	4.00E-06	3.51E-01	9.01E-03
0.5	6	4	1.09E-05	1.84E-03	0.000384	1.39E-04	2.43E-01	0.003516
0.5	6	5	1E-06	0.000847	0.000132	2.60E-05	2.01E-01	3.28E-03
0.5	10	3	3.36E-05	0.001498	0.000361	2.30E-05	0.194922	2.51E-03
0.5	10	4	1.21E-06	4.58E-04	9.41E-05	9.61E-06	1.46E-01	1.87E-03
0.5	10	5	1.61E-04	6.08E-05	1.85E-05	1.60E-07	1.14E-01	0.000841
0.75	2	3	3.88E-04	2.27E-02	5.45E-03	0.001616	1.10E+00	7.47E-02
0.75	2	4	1.44E-06	0.009683	0.001673	0.000324	0.755161	0.027956
0.75	2	5	2.16E-04	7.96E-03	2.01E-03	6.89E-05	5.89E-01	1.69E-02
0.75	6	3	1.00E-06	1.73E-03	2.59E-04	4.10E-05	3.34E-01	6.58E-03
0.75	6	4	1.85E-05	5.71E-04	4.76E-05	4.00E-08	2.40E-01	3.49E-03
0.75	6	5	1.96E-06	4.20E-04	5.33E-05	1.44E-05	1.96E-01	0.00184
0.75	10	3	1.44E-06	0.000807	0.000139	1.52E-05	1.92E-01	2.10E-03
0.75	10	4	1.94E-05	5.24E-04	1.42E-04	2.50E-07	1.45E-01	1.39E-03
0.75	10	5	1.76E-05	4.41E-04	1.19E-04	3.60E-07	1.12E-01	0.000524
0.9	2	3	3.28E-04	1.69E-02	0.003956	1.38E-03	1.007213	0.05598
0.9	2	4	4.41E-06	0.007225	0.001204	0.000279	6.94E-01	0.019768
0.9	2	5	3.25E-05	4.50E-03	0.00087	1.96E-06	0.544644	0.012122
0.9	6	3	1.00E-06	0.00131	0.00021	6.56E-05	3.17E-01	4.61E-03
0.9	6	4	1.60E-07	6.97E-04	0.000112	1.09E-05	2.31E-01	0.00207
0.9	6	5	1.00E-06	3.72E-04	6.4E-05	5.78E-05	1.88E-01	0.001043
0.9	10	3	2.89E-06	7.56E-04	0.000135	3.14E-05	1.84E-01	1.52E-03
0.9	10	4	3.60E-07	0.000286	4.62E-05	2.12E-05	1.41E-01	9.00E-04
0.9	10	5	1.00E-08	2.13E-04	2.81E-05	4.00E-06	1.09E-01	0.000404
1	2	3	5.33E-05	1.26E-02	2.37E-03	0.001253	0.840706	0.033526
1	2	4	8.1E-07	0.004277	0.000581	0.00024	5.57E-01	8.74E-03
1	2	5	1.21E-06	2.13E-03	2.86E-04	3.36E-05	4.26E-01	0.004396
1	6	3	4.90E-05	1.99E-03	4.62E-04	1.30E-05	2.86E-01	2.98E-03
1	6	4	7.29E-06	3.42E-04	2.92E-05	7.84E-06	1.97E-01	9.12E-04
1	6	5	5.76E-06	4.49E-04	8.84E-05	2.12E-05	1.57E-01	3.57E-04
1	10	3	3.61E-06	0.000543	0.000104	3.60E-05	1.77E-01	9.30E-04
1	10	4	1.30E-04	5.06E-04	2.43E-04	6.76E-06	1.27E-01	4.33E-04
1	10	5	2.56E-06	9.22E-05	2.03E-05	1.44E-06	9.87E-02	1.56E-04

Table 4.12 Biases of AM,GM, and HM for gamma and lognormal distribution from simulation process by using original Dell & Clutter model.

ρ	c	h	Gamma (4,2)			LogNormal (2.9761, 0.1980)		
			AM	GM	HM	AM	GM	HM
0.5	2	3	0.000202	0.143489	0.021904	3.97E-05	1.19E-02	0.002601
0.5	2	4	2.92E-05	9.07E-02	1.12E-02	3.02E-05	5.99E-03	1.27E-03
0.5	2	5	1E-08	0.057504	0.007006	2.25E-04	2.21E-03	2.43E-04
0.5	6	3	1.44E-06	2.10E-02	2.54E-03	1.96E-06	1.43E-03	3.24E-04
0.5	6	4	5.04E-05	1.26E-02	0.001815	9.02E-05	4.24E-04	3.36E-05
0.5	6	5	1.52E-05	0.006162	0.000571	2.56E-06	5.29E-04	1.14E-04
0.5	10	3	0.000137	0.004277	0.000185	6.40E-07	0.000488	1.14E-04
0.5	10	4	8.65E-05	5.52E-03	8.76E-04	1.30E-04	7.45E-04	3.76E-04
0.5	10	5	2.12E-05	3.04E-03	4.45E-04	3.61E-06	1.39E-04	2.4E-05
0.75	2	3	6.92E-04	1.13E-01	1.79E-02	4.1E-05	1.10E-02	3.04E-03
0.75	2	4	0.001884	0.063655	0.011599	2.5E-05	0.00402	0.000829
0.75	2	5	1.24E-03	4.08E-02	6.91E-03	1.23E-05	2.38E-03	6.45E-04
0.75	6	3	1.76E-05	1.55E-02	1.48E-03	5.63E-05	1.54E-03	5.34E-04
0.75	6	4	0.001354	7.43E-03	0.002372	4.90E-07	5.06E-04	1.21E-04
0.75	6	5	1.90E-04	2.43E-03	4.49E-04	2.50E-07	4.97E-04	0.000128
0.75	10	3	0.000524	0.002663	0.000718	2.89E-06	4.16E-04	1.23E-04
0.75	10	4	3.50E-04	1.18E-03	3.61E-04	8.65E-05	4.37E-04	2.31E-04
0.75	10	5	1.14E-03	7.34E-04	6.55E-04	3.60E-07	9.60E-05	2.12E-05
0.9	2	3	7.40E-05	9.63E-02	0.011257	2.12E-05	0.007691	0.002061
0.9	2	4	4.90E-05	0.053361	0.005791	4E-06	3.28E-03	0.000841
0.9	2	5	3.60E-07	3.42E-02	0.003102	2.25E-06	0.00131	0.00032
0.9	6	3	2.12E-05	0.015178	0.001332	7.4E-05	1.36E-03	5.02E-04
0.9	6	4	9.80E-05	7.48E-03	0.000936	1.96E-06	3.13E-04	6.72E-05
0.9	6	5	1.60E-07	3.88E-03	0.00031	3.60E-07	2.82E-04	6.4E-05
0.9	10	3	1.82E-04	3.97E-03	9.22E-05	1.44E-06	2.46E-04	5.33E-05
0.9	10	4	1.96E-06	0.001875	0.000112	1.61E-04	5.02E-04	3.10E-04
0.9	10	5	8.84E-05	2.47E-03	0.000392	7.29E-06	1.14E-04	4.49E-05
1	2	3	1.25E-04	7.89E-02	8.84E-03	3.14E-05	0.005806	0.001624
1	2	4	2.3E-05	0.033893	0.002894	1.69E-06	1.77E-03	3.96E-04
1	2	5	2.60E-05	1.94E-02	1.16E-03	1.02E-05	8.24E-04	0.000228
1	6	3	4.41E-06	1.10E-02	8.76E-04	3.97E-05	9.30E-04	3.28E-04
1	6	4	8.46E-05	5.00E-03	5.66E-04	1.16E-05	1.42E-04	1.85E-05
1	6	5	1.60E-07	1.91E-03	9.60E-05	1.21E-06	1.74E-04	4.90E-05
1	10	3	0.00011	0.003387	9.03E-05	1.00E-06	2.53E-04	7.22E-05
1	10	4	4.41E-06	1.41E-03	7.74E-05	1.30E-04	3.50E-04	2.28E-04
1	10	5	3.48E-05	1.66E-03	1.90E-04	2.56E-06	5.04E-05	1.94E-05

CHAPTER FIVE

CONCLUSION

In this study, the proposed estimator of harmonic mean and geometric mean of parameter of population are evaluated by computer simulation and the conclusions are following:

1. The parameters HM and GM of population can be estimated by using Equation 3.6 and 3.7.
2. Based on the result of simulation, estimation of HM and GM by using RSS gives better efficiency than SRS.
3. The correlation coefficient (ρ) gives effect to RE. If the value of ρ is increased the values of RE increases as well, except for a few data sets from exponential distribution.
4. The set size of sample can increase the efficiency of RSS, except in the case of HM in the exponential distribution.
5. The relationship of relative efficiency between AM, GM, and HM when the data sets are generated from normal distribution is as follows

$$RE_{HM} \leq RE_{GM} \leq RE_{AM}$$

6. The mean estimators of HM and GM based on RSS are unbiased estimator, when the data sets are generated from normal, lognormal, gamma, and exponential distribution.
7. According to simulation results from normal distribution, the coefficient of correlation from the imperfect judgment ranking model of Dell & Clutter (1972) and Adam & Levine (1972) yields the greater value of relative efficiency than the value from the correlation coefficient model of Dell & Clutter (1972) with modification based on regression concept.

In addition, there are two problems which can be the topics for discussion or future research. First problem is the open problem about the best method of coefficient of correlation for measuring the perfectness of correlation between response data and concomitant variable. Second is about the variance of estimator which can be an additional tool for estimating the parameter of response data.

REFERENCES

Burk, F. (1985). By all means. *The American Mathematical Monthly, 92*(1), 50.

Chen, Z., Bai, Z., & Sinha, B. K. (2003). *Ranked set sampling, theory and applications*. New York: Springer.

Cochran, W. G. (1977). *Sampling techniques*. Massachusetts: John Wiley & Son.

David, H. A., & Levine, D. N. (1972). Ranked set sampling in the presence of judgment error. *Biometrics*, *28*, 553–555.

Dell, T. R., Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, *28*, 545-555.

Ferger, W. F. (1931). The nature and use of the harmonic mean. *Journal of American Statistics Association, 26* (173), 36-40.

Francis, A. (2004). *Business mathematics and statistics*. Croatia: Cengage Learning.

Gomez-Hernandez, J. J., & Gorelick, S. M. (1989). Effective ground-water model parameter values: Influence of spatial variability of hy-draulic conductivity, leakance, and recharge. *Water Resource Repository*, *25*(3), 405–419.

Ibrahim, K., Syam, M., & Alomari, A. I. (2010). Estimating the population mean using stratified median ranked set sampling. *Applied Mathematical Sciences*, *4*(47), 2341-2354.

Limbrunner, J. F., Vogel, R. M., & Brown, L. C. (2000). Estimation of harmonic mean of a lognormal variable. *Journal of Hydrologic Engineering*, *5*(1), 59-66.

Lohr, S. L. (1999). *Sampling: design and analysis*. Arizona: Duxbury Press.

Misra, S., & Gupta, R. K. (2010).Estimation of geometric mean of skewed populations with and without using auxiliary information. *Journal of Probability and Statistical Studies*, *3*(2), 35-46.

Ozturk, O., & Balakrishnan, N. (2009). Exact two-sample nonparametric test for quantile difference between two populations based on ranked set samples. *Annual Institute of Statistics Mathematics*, *61*, 235–249.

Rossman, L. A. (1990). Design stream flows based on harmonic means. *Journal of Hydrology Engineering*, *ASCE, 116*(7), 946–950.

Saatci, A. M. (1989). Harmonic mean conductivity in declining rate filters. *Journal of Environmental Engineering*, *ASCE, 115*(2), 462–466.

Scheaffer, R. L., Mendenhall. W., & Ott. L. (1990). *Elementary survey sampling*. California: Duxbury Press.

Spizman, L. (2008). A note on utilizing the geometric mean: when, why and how the forensic economist should employ the geometric mean. *Journal of Legal Economics*, *15*(1), 43-55.

Wolfe, D. N. (2012). Ranked set sampling; its relevance and impact on statictical inference. *International Scholarly Research Network*. Retrieved April 14, 2014, from http://www.hindawi.com/journals/isrn/2012/568385/.

APPENDIX

Matlab code;

{SRS}

```
mu = [200 20];
SIGMA = [100 72; 72 64];
data = mvnrnd(mu,SIGMA,10000);
gen=transpose([1:10000]);
for r=(1:10000)
   % taking random data
   s=randsample(gen,150);
   for m=(1:150)
      zxx(m)= data(s(m),1);
      zyy(m)= data(s(m),2);
   end
end
```

{RSS}

```
mu = [200 20];
SIGMA = [100 72; 72 64];
data = mvnrnd(mu,SIGMA,10000);
gen=transpose([1:10000]);
for r=(1:10000)
   % taking random data
```

```
s=randsample(gen,150);
for m=(1:150)
    zxx(m)= data(s(m),1);
    zyy(m)= data(s(m),2);
end
%grouping in cycle
for l=(1:6)
    zx4(1:25,l)= zxx([1+25*l-25:25+25*l-25]);
    zy4(1:25,l)= zyy([1+25*l-25:25+25*l-25]);
end
%grouping in rank
for n=(1:6)
    for o=(1:5)
      % zx5(1:5,o,n)= zx4([1+5*o-5:5+5*o-5],n);
      % zy5(1:5,o,n)= zy4([1+5*o-5:5+5*o-5],n);
       zx6(1:5,o)= zx4([1+5*o-5:5+5*o-5],n);
       zy6(1:5,o)= zy4([1+5*o-5:5+5*o-5],n);
    end
%sorting
for q=(1:5)
     for i = (1:5-1)
```

```
            minElem = zx6(i,q);
            minIndex = i;
              %This for loop can be vectorized, but there will be no significant
                %increase in sorting efficiency.
          for j = (i:5)
            if zx6(j,q) <= minElem
            minElem = zx6(j,q);
            minIndex = j;
            end
          end
        if i ~= minIndex
        zx6([minIndex i],q) = zx6([i minIndex],q); %Swap
        zy6([minIndex i],q) = zy6([i minIndex],q);
        end
    end
end
%sorting end
    for p=(1:5)
    zx8(p,n,r)=zx6(p,p);
    zy8(p,n,r)=zy6(p,p);
    end
 end
```

```
 %for converting data and calculate mean, HM, and GM
 c=[1];
 for b=(1:6)
     for a=(1:5)
         zx9(c,r)= zx8(a,b,r);
         zy9(c,r)= zy8(a,b,r);
         c=c+1;
     end
 end
% meanarx(r) = mean(zx9);
 %meanharx(r) = harmmean(zx9);
 %meangeox(r) = geomean(zx9);
 % meanary(r) = mean(zy9);
 %meanhary(r) = harmmean(zy9);
 %meangeoy(r) = geomean(zy9);
end
{ calculating standard deviation}
sdofmmx=0;
sdofmmx2=0;
sdofmhx=0;
  sdofmhx2 = 0;
  sdofmgx =0;
```

```
    sdofmgx2 = 0;


    sdofmmy =0;
    sdofmmy2 = 0;
    sdofmhy =0;
    sdofmhy2 = 0;
    sdofmgy =0;
    sdofmgy2 = 0;
for d=(1:10000)
    sdofmmx(d)=((meanx(d)-mmx)^2)/(10000-1);
    sdofmmx2  = sdofmmx2+sdofmmx(d);
    sdofmhx(d)=((meanhx(d)-mhx)^2)/(10000-1);
    sdofmhx2 = sdofmhx2+sdofmhx(d);
    sdofmgx(d)=((meangx(d)-mgx)^2)/(10000-1);
    sdofmgx2 = sdofmgx2+sdofmgx(d);


    sdofmmy(d)=((meany(d)-mmy)^2)/(10000-1);
    sdofmmy2 = sdofmmy2+sdofmmy(d);
    sdofmhy(d)=((meanhy(d)-mhy)^2)/(10000-1);
    sdofmhy2 = sdofmhy2+sdofmhy(d);
    sdofmgy(d)=((meangy(d)-mgy)^2)/(10000-1);
    sdofmgy2 = sdofmgy2+sdofmgy(d);
```

```
end


sdofmmx2
   sdofmhx2
   sdofmgx2
   sdofmmy2
   sdofmhy2
   sdofmgy2
{Generating normal data}
R = normrnd(20,4,[1 10000]);
R= transpose(R);
eror= normrnd(0,3.464,[1 10000]);
eror= transpose(eror);
X = R-eror;
data(:,1)=X;
data(:,2)=R;
{ Generating gamma distribution}
Y = gamrnd(4,2,[1 10000]);
Y = transpose(Y);


eror= normrnd(0,2.6458,[1 10000]);
```

```
eror= transpose(eror);

X = Y-eror;

X= X(:,1).^2;

X= sqrt(X);

data(:,1)=X;

data(:,2)=Y;

{ Generating exponential data}

R = exprnd(4,[1 10000]);

R= transpose(R);

eror= normrnd(0,2.6458,[1 10000]);

eror= transpose(eror);

X = R-eror;

X= X(:,1).^2;

X= sqrt(X);

data(:,1)=X;

data(:,2)=R;
```