USING SIMULATION TO STUDY

THE SAMPLING DISTRIBUTION OF SOME

WELL KNOWN STATISTICS

BY

HUDA IBRAHIM RAMMAHA

A Thesis

Presented to the Graduate School

Birzeit University

In Candidacy of Master of Science

In Applied Statistics

Birzeit University

Birzeit University 2004

CERTIFICARE OF APPROVAL

Approved and recommended for acceptance as a thesis in partial fulfillment

of the requirements for the degree of Master of Science.

……………………………………
(Date)

Accepted………………………….
(Date)

Special committee directing

The thesis of Huda Rammaha.

…………………

…………….

Chairman: Dr. Elias Dabeet.

…………………

……………..

Dr. Othman Abu Libdeh.

………………

………………

Dr. Simon El
Araj.

ACKNOWLEDGMENTS

I wish to acknowledge all the many people who have contributed to the writing of this thesis. In particular, I would express my thanks to the following people:

My advisors, Dr. Elias Dabeet, Dr. Othman Abu Libdih and Dr. Simon El Araj of Birzeit University, for their encouragement and constant guidance which contributed significantly to the accuracy and clarity of my thesis.

An acknowledgement of great thanks goes to my husband for his aid, interest, constant encouragement and understanding.

DEDICATION

To my husband

Mr. Mohammad Mashni

And

To my three lovely children

Issam, Duaa and Lynn Mashni

Table of Contents

Table of Tables

Table of figures

Table of appendices

Appendix 1

Appendix 2

distribution of the sample median. EXP (2).

Appendix 3

Table of symbols	
SYMBOL	**MEANING**
n	Sample size
μ	Population mean
θ	Population parameter
$\hat{\theta}$	Estimation of θ
$\bar{x}$	Sample mean
$\mu_{\bar{x}}$	Mean of the sample mean
$\mathring{x}$	Sample median
$\mu_{\mathring{x}}$	Mean of the sample median
σ^2	Population variance
σ	Population standard deviation
s^2	Sample variance
s	Sample standard deviation
$\sigma_{\bar{x}}$	Standard deviation of the sample mean
$\sigma_{\mathring{x}}$	Standard deviation of median
ν	Degrees of freedom

Definition of terms

A population: The set of all the elements of interest [16].

A sample: Is a subset of the population [15].

A parameter: is a measurement on a population that characterizes one of its features [16].

A point estimator: is a sample statistic that predicts the value of the corresponding parameter [15].

A statistics: A statistic is a measure on the items in a random sample. An example of a statistic is the mean (i.e. average) of the measures in the sample [15].

A sampling distribution: The Sampling Distribution of a statistic is the set of values that we would obtain if we draw an infinite number of random samples of size n from a given population and calculate the statistic on each sample. [16].

Unbiasedness: sampling distribution of the point estimator is centered on the population parameter

$E(\hat{\theta}) = \theta$ [16]

Efficient: the point estimator has the smallest possible standard deviation of all similar point estimators (i.e. $Var(\hat{\theta})$ is smallest). [16]

Consistent: the point estimator tends toward a population parameter as the sample size increases (i.e. $E(\hat{\theta}) \to \theta$ and $Var(\hat{\theta}) \to 0$ as $n \to \infty$). [16]

Simulation: is a numerical technique for conducting experiments on the computer [12].

Simulation in statistical terminology: simulation is an artificial data generation process, driven by model design and parameter settings. The output is a sample, and drawing conclusions about how the process might change if the conditions or parameters underlying the process are changed [5]. Sample statistics are computed and tested as estimators of the population parameters.

Probabilistic simulation: use sample generated from known underlying distributions [20].

Bootstrap simulation method: Bootstrap method can be used to estimate the population parameter when the distribution of the population is unknown by creating a simulated original population by repeating sample from the original sample with replacement [20]. The basic idea behind bootstrapping is that if the sample is a good representative of the population, the sampling distribution of interest may be estimated by generating a large number of new samples (called resample) from the original sample. In another way, bootstrap treats the sample as if it is the population [18]. The only assumption of the bootstrap method is that the original sample is representative of the population from which it was drawn [7].

The normal distribution: is a continuous distribution with probability density function:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} \qquad -\infty < y < \infty \text{ [13].}$$

The parameters μ and σ^2 are the mean and the variance respectively, of the normal random variable y.

The uniform distribution: is a continuous distribution with the density function:

$f(y) = \frac{1}{b-a}$ if $a \le y \le b$

0 elsewhere

$\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(b-a)^2}{12}$ [13].

The exponential distribution: is a continuous probability distribution with the probability density function.

$$f(y) = \frac{e^{-y/\beta}}{\beta} \qquad (0 \le y \le \infty)$$

with $\mu = \beta$ and $\sigma^2 = \beta^2$ [13].

ملخـص الرسالة

:الهدف من هذه الرسالة هو استخدام طريقتين من طرق محاكاة الحاسوب الإحصائية وهما

لدراسة الصفات الرئيسية الثلاثة وهي: المركز، (bootstrap) و (probabilistic)

الانتشار والشكل. والتي تعتبر ركائز لتوزيع المعاينة لمعدّل العينة، ووسـيط العينـة وتبـاين

. العينة. وذلك حسب متغيرات أساسية كحجم العينة وتوزيع المجتمع الذي أخذت منه العينة

.وقد تم استخدام أساليب متعددة في الدراسة منها الرسم والتحليل

-:ويمن تلخيص نتائج المحاكاة لهذه الدراسة على النحو التالي

1- إذا كان حجم العينة كبيـــــراً فإن توزيع المعاينة للوسط الحسابي للعينة ووسيط العينـــة وتباين العينة يتبع التوزيع الطبيعي مع قليل من التحيز. بغض النظر عن توزيع المجتمع الذي أخذت منه العينة.

2- إذا كان حجم العينة صغيراً فإن هذه المقدرات التي تمّ ذكرها أعلاه تتبع التوزيع الطبيعي اعتماداً على توزيع المجتمع الذي أخذت منه العينات.

3- شكل التوزيع للكمية الإحصائية $\frac{(n-1)s_i^2}{\sigma^2}$ تتبع لتوزيع chi-square مع (n-1) درجات حريـــــة. وذلك إذا كان حجم العينة صغيراً وتوزيع المجتمع يتبع التوزيع الطبيعي.

4- الوسيط الحسابي للعينة وتباين العينة غير مرتبطتين إحصائياً مع بعضهما البعض. إذا كان توزيع المجتمع يتبع التوزيع الطبيعي أو التوزيع المنتظم. ويكون هذان المقدران مرتبطان إذا كان توزيع المجتمع يتبع التوزيع الاسّي.

Abstract

The focus of this study is to use the probabilistic and the bootstrap simulation methods to investigate graphically, computationally and analytically the three main characteristics: location, spread and the shape of the sampling distribution of the three well known sample statistics: the sample mean, the sample median, and the

sample variance by varying the sample size and distribution of the underlying population.

The simulated results of this study can be summarized as follows:

a) The sampling distribution of the sample mean, the sample median and the sample variance can be approximated by a normal distribution, with a very small bias, provided the sample size is large. However, for small sample size, the normality of the sampling distribution of these statistics depends strongly on the underlying population distribution. The simulated results have showed that the sample mean is less variance than the sample median.

b) The graphical and analytical simulated results indicated that the shape of the sampling distribution of the statistics $\frac{(n-1)s_i^2}{\sigma^2}$ is approximately chi-square with n-1 degrees of freedom for small sample size and the underlying distribution is normal. Not symmetric for small degrees of freedom.

c) The simulated results showed that the sample mean and the sample variance are not correlated when the underlying population distributions are uniform or normal; however, the simulated results showed that these statistics are correlated in the exponential case.

Chapter One

1.1 Introduction

The Probabilistic and Bootstrapping simulation methods offer ways to study the sampling distributions of statistics graphically through histograms, and to visualize the distribution of data sets.

Probabilistic simulations were used to study the sampling distribution of the sample mean $\overline{x}$ by many researchers Weir, McManus, and Kiely Marasinghe et al., delMas et al.[23], they have studied the sampling distribution and the central limit theorem by specifying and changing the population underlying distribution with different sample sizes.

Arnholt [12], Schwarz and Sutherland [21] and Hesterberg [19], have used probabilistic simulation to visualize and illustrate many properties of the sampling distributions that depend on the specific statistic, sample size, and the population distribution. Also Ng and Wong [14] have demonstrated the central limit theorem graphically by choosing a specific distribution from which the data are generated, a sample size and the number of samples to be drawn.

In this thesis, we have used two methods of simulations, the Probabilistic and the Bootstrapping simulation methods to study the sampling distribution of the sample mean, sample median and sample variance graphically and analytically.

To be specific, we have studied the behavior of the sampling distribution of these statistics computed from 1000 repeated samples with replacement of different sizes (i.e. n=2, 5, 10, 15, 20, 30, 50 and 100) drawn from the symmetric normal distribution, the symmetric uniform distribution and the non-symmetric exponential distribution in order to assess the effect of symmetry on the sampling distribution of the above statistics. This simulation study have allowed us to gain insight into the behavior of the statistics: sample mean, sample median and sample variance by comparing the theoretical results with the simulated results. In particular, we have compared the center, the spread and the shape of the simulated results with the theoretical results.

On the other hand, we have studied the correlation between the sample mean and sample variance.

In general we have illustrated four very important theorems in statistics using computer simulation.

1) The law of large samples states that as sample size increases, the sample mean is likely to be more accurate estimate of the true population (i.e. $\mu = \mu_{\bar{x}}$).

2) The central limit theorem states that as sample size get larger; the sampling distribution of the sample mean becomes approximately normal, regardless of the shape of the underlying population distribution.

3) The median theorem states that if the population is $N\left(\mu,\sigma^2\right)$ then the medians of random samples of size n are distributed with a mean of the medians $\mu_{\tilde{x}}$ and with standard deviation which is equal to the population standard deviation divided by the square root of the sample size and multiplied by the constant 1.25 (i.e. $\sigma_{\tilde{x}} = 1.25\frac{\sigma}{\sqrt{n}}$).

4) The independence of sample mean and the sample variance states that if we draw a random sample from N (0, 1), then the sample mean and the sample variance are not correlated.

Through Probabilistic and Bootstrapping simulation, graphically, analytically and computationally, we have answered the following main questions.

1) Are the sampling distribution of the sample mean and sample median unbiased estimators of the true population mean, for the above mentioned probability distributions and for each sample size?

2) Are the standard deviations of the sample mean and the sample medians equal to the population standard deviation divided by square root of the sample size, for the above mentioned probability distributions and for each sample size?

3) What is the overall shape of the sampling distribution graphically and analytically using Kolmogorov-smirnov goodness of fit test, for the above mentioned probability distributions and for each sample size?

4) Is the sampling distribution of the quantity $\frac{(n-1)s_i^2}{\sigma^2}$ follows a chi-square with (n-1) degrees of freedom graphically and analytically using

Kolmogorov-smirnov goodness of fit test, for the above mentioned probability distribution and for each sample size?

5) Are the sample mean and the sample variance uncorrelated regardless of the underlying population distribution?

1.2 Simulation and sampling distribution:

In this study we have used Probabilistic and Bootstrap simulation to study the sampling distribution by drawing 1000 samples of size “n” from a given population and computing the sample mean, sample median, and the sample variance for each sample, then comparing our simulated results for the sampling distribution for each of the above statistics with the theoretical results, and checked if these results are characterized with the properties of a good estimator to include unbiasedness (center), consistency (spread) and the shape.

Every simulated statistic has a sampling distribution. For example, the distribution of the means of an infinite number of samples would be called the sampling distribution of the mean. For the distribution of an infinite number of sample median would be called the sampling distribution of the median, similarly for the sample variance.

We have used the probabilistic and the bootstrap simulation to compare the theoretical results with the simulated results graphically and analytically using the Kolmogorov-Smirnov goodness of fit test that can be used to test the goodness of fit to any continuous probability distribution. Previous studies utilized only the probabilistic simulation method and the graphical results only.

1.2.1 Simulation of the sampling distribution of the mean:

In this study, the Probabilistic and Bootstrap simulation methods have been used to study graphically, computationally and analytically the three important properties of the sampling distribution of the sample mean by simulating different sample sizes from three different populations: N(0,1), U(10,20), EXP(2).

The expected value of the sampling distribution of the mean is represented by the symbol $\mu_{\bar{x}}$; a sampling distribution of the sample mean may also be described with a parameter corresponding to the standard deviation, symbolized by $\sigma_{\bar{x}}$ [3].

We have used the two methods of simulation to verify graphically and analytically if these parameters are closely related to the parameters of the population distribution, with the relationship being described by the Central Limit Theorem.

The Central Limit Theorem essentially states "that the mean of the sampling distribution of the mean ($\mu_{\bar{x}}$) equals the mean of the population (μ) and that the standard error of the mean ($\sigma_{\bar{x}}$) equals the standard deviation of the population (σ) divided by the square root of n as the sample size gets infinitely larger (n $\longrightarrow \infty$)". In addition, "the sampling distribution of the mean will approach a normal distribution" [6].

These relationships can be summarized as:

1) $\mu = \mu_{\bar{x}}$ (Unbiasedness)

2) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (Consistency)

3) The sampling distribution of the sample mean will approach a normal distribution as n $\longrightarrow \infty$ [8].

In practice, an infinite sample size is not practical. The Central Limit Theorem is very powerful. In most situations, this theorem works reasonably well with n greater than 10 or 20. Thus, it is possible to closely approximate what the distribution of sample means looks like, even with relatively small sample sizes [6].

In general the purpose of the central limit theorem simulation is to demonstrate that under certain conditions and for large enough samples, the sampling distribution of the sample mean can be approximated by normal distribution. The conditions under these properties can be illustrated by direct manipulation of sample size, shape and symmetry of the underlying distribution and number of samples drawn [9].

1.2.2 Simulation of the sampling distribution of the median:

The probabilistic and bootstrap simulation methods have been used to study graphically, computationally and analytically the three important properties of the sampling distribution of the sample median by simulating different sample sizes from three different populations: N(0,1), U(10,20), EXP(2).

If the population follows $N\left(\mu,\sigma^2\right)$, the medians of random samples of size n are distributed with a mean of the median denoted by $\mu_{\tilde{x}}$ and a standard deviation denoted by $\sigma_{\tilde{x}}$ [3]

We have used the two methods of simulation to verify, graphically and analytically, if these parameters are closely related to the parameters of the population distribution.

These relationships can be summarized as:

1) $\mu = \mu_{\hat{x}}$ (Unbiasedness).

2) $\sigma_{\hat{x}} = 1.25 \frac{\sigma}{\sqrt{n}}$ (Consistency).

3) The distribution of sample medians is nearly normal if n is large (Shape) [3].

1.2.3 Simulation of the sampling distribution of the sample variance:

The Probabilistic and Bootstrap simulation methods have been used to study graphically and analytically the properties of the sampling distribution of the sample variance by simulating different sample sizes from three different populations: N(0,1), U(10,20), EXP(2).

Independent samples of size n have been simulated from a normal distribution with known variance σ^2, for each sample the sample variances s_1^2, s_2^2, ... s_k^2, ... are computed, then we calculate for each sample *i* the quantity

$$w_i = \frac{(n-1)s_i^2}{\sigma^2}$$

The histogram of these quantities would converge in distribution to the chi square distribution (χ^2) with n-1 degrees of freedom.

1.2.4 Simulation of the independence of the sample means and sample variance:

The Probabilistic and the Bootstrap simulation methods were used to study analytically the correlation between the sample variance the sample mean by simulating different sample sizes from three different populations: N(0,1), U(10,20), EXP(2).

We have used the two methods of simulation to verify analytically that the quantities v_i and w_i that shown below are uncorrelated when the underlying distribution is $N(\mu, \sigma^2)$

$$w_i = \frac{(n-1)s_i^2}{\sigma^2} \qquad \text{And} \qquad v_i = \sqrt{n}\,\overline{x_i}$$

Chapter Two

Literature review

Simulation has had a major impact on the practice of statistics and understanding statistical methods and concepts. The central limit theorem was one of the more popular topics that were illustrated by simulation. Also using simulation to understand the concepts related to students-t distribution using Minitab.

In addition, simulation was used to illustrate the idea of inference and sampling error and generating a sample to find a confidence interval. Moreover, the binomial distribution and regression analysis are other topics that were studied by using simulation.

Simulation was used to explore the properties of any sampling distribution and to understand the sampling distribution and the central limit theorem by specifying and changing the shape of a population, choosing different sampling sizes, and exploring sampling distributions by randomly drawing large numbers of samples [10].

Many statistics educators have studied sampling distributions and the central limit theorem. For examples Ng and Wong have used simulation experiments on the Internet to illustrate Central Limit Theorem (CLT) concepts. The CLT can be demonstrated graphically, the program begins by allowing the user to choose a distribution, from which the data are to be generated, a sample size for the sampling distribution for the mean, and the number of samples to be drawn. By changing the sample size, the user can observe how fast the probability histogram approaches the normal curve as the sample size increases. The program also allows the user to compare sampling distributions of other statistics as well such as the median and standard deviation [14].

Many other statistics educators have used simulation exercises on the Internet for the CLT like West and Ogden [24] and with other topics by Schwarz and Sutherland [21].

Kersten [11] have used Probabilistic simulation methods to clarify concepts and theorems of statistics (such as the CLT) the author conducted three simulations, each of 300 random samples for $n = 1$, $n = 4$, and $n = 16$ from the IRANDOM population. The mean of each sample was computed and the resulting means were used to construct a histogram. This will help us see how mean of the sample means is close to the population mean but the standard deviation decrease for the 300 random samples as n increases. We can also see from the histograms that as n becomes larger, the distribution of the sample means becomes more similar to the normal distribution.

Dambolena [4] have used computer simulations to understand the CLT. Using BASIC® programming, he suggested drawing a random sample of size 30 from a discrete uniform population with mean μ and standard deviation σ, computing the sample mean, and repeating this procedure 1000 times. Using MINITAB®, the 1000 means obtained from samples of size $n = 30$ can be output in a separate file that would subsequently be used to generate histograms and to illustrate the concepts of the CLT.

Yu, Behrens, and Anthony [4] have used simulation to gain a better understanding of the concept of an expected value, the shape of distributions of varying sizes, and the meaning of a sampling distribution.

Yu et al. [25] have concluded that some aspects of the CLT can be clearly illustrated by computer software but some cannot. For example, simulation can perform the function of showing the process of a sampling distribution, but abstract concepts such as equality, independence, and the relationship between the CLT and hypothesis testing are difficult to present.

Weir, McManus, and Kiely, Marasinghe et al., delMas et al. [23] have used simulation to study the sampling distributions and the CLT by specifying and changing the shape of a population, choosing different sample sizes, and exploring sampling distributions by randomly drawing large numbers of samples.

Arnholt [2] have used simulation to study the concept of a sampling distribution, which can provide greater insight and a more thorough understanding of statistics and their distributions. They have proved that every statistic has a sampling distribution and using simulation, can provide a concrete way to illustrate this as well as a way to reveal how other factors, such as sample size, and affect the sampling distribution. Sampling distributions can be generated for many commonly used statistics (such as the sample sum, mean, median, standard deviation, variance, range, and *t*-statistic). Visualizations can illustrate many of the properties mentioned previously including the facts that different statistics have different sampling distributions that depend on the specific statistic, sample size, and the underlying distribution, that the variability in the sampling distribution can be decreased by increasing the sample size, and that for large samples, the sampling distribution can be approximated by a normal distribution.

Tim C. Hesterberg [19] had used bootstrapping as reinforcement of providing the experience in statistical concepts. He has used to estimate sampling distribution, when they have data but do not know the underlying distribution.

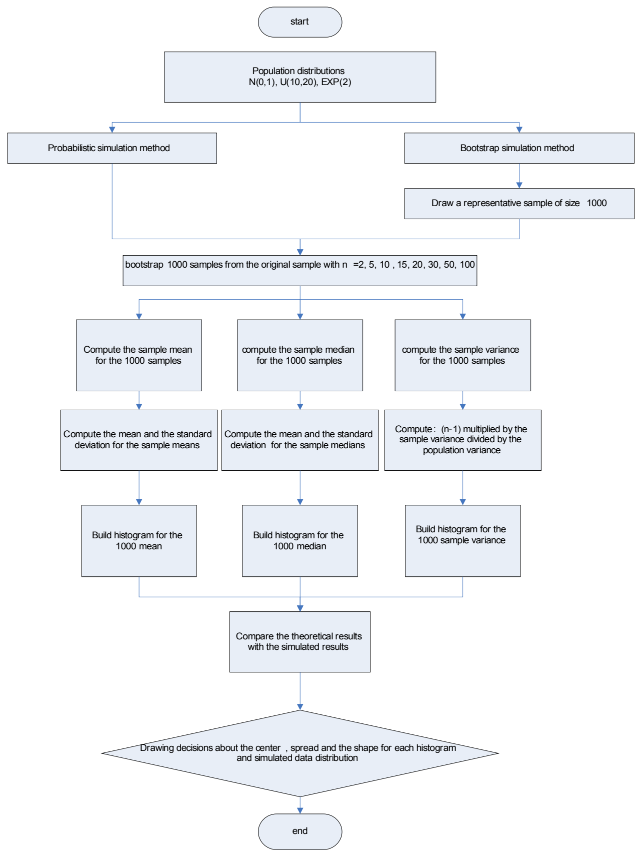
Chapter Three

Simulation procedure and Results

In this chapter we have used the Probabilistic and the Bootstrap simulation methods to study the sampling distribution of the sample mean, the sample median and the sample variance in which the samples are generated from the known underlying population distribution and the Bootstrap method in which the sample is a good approximation of the population, the sampling distribution of interest may be estimated by generating a large number of new samples (called resample) from the original population. Two symmetric and one non-symmetric population distributions were used in this study. These population distributions are the normal, the uniform and the exponential distribution. These populations were selected in order to assess the effect of symmetry and normality on the sampling distribution of the above mentioned statistics.

The following flowchart (1) summarizes the procedure:

Flowchart (no. 1)



3.1 simulation procedure and Sampling distribution of the sample mean:

In this section we have studied graphically, computationally and analytically the sampling distribution of the sample mean using two methods of simulation, the Probabilistic method and the Bootstrap method.

Without loss of generality and for simulation purposes, Minitab, SPSS, Excel and Statistica softwares were used in this thesis to generate random samples from the three distributions with their specific parameter values.

1) Normal distribution with mean 0 and variance 1, N (0, 1).

2) Uniform distribution with lower limit 10 and upper limit 20, U (10, 20).

3) Exponential distribution with mean 2, EXP (2).

Simulation procedure

1000 samples were simulated from the above mentioned distributions with sample sizes n=2, 5, 10, 15, 20, 30, 50 and 100. Then histograms were drawn to explore the shape of the sampling distribution of the sample mean with parameters $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.

Kolmogorov-smirnov goodness of fit test for normality was used in each case to assess the fit.

Using simulations and in particular the histograms in appendix 1, the tables (3.1.1-3.1.6) and the bar-charts (1.1-1.3), we arrive at the following conclusions:

1) The mean of the sample means (simulated mean of the means) $\mu_{\bar{x}}$ from normal, uniform and exponential distribution are approximately equal to the population mean regardless of the sample size. This agrees with the theory that the sample mean is an unbiased estimator of the population mean regardless of the underlying population distribution (i.e. $\mu = \mu_{\bar{x}}$; unbiasedness property).

2) Simulated Standard deviation of the means is approximately equal to the population standard deviation times $\frac{1}{\sqrt{n}}$ when samples were drawn from normal, uniform and exponential for all sample sizes. This is as expected from theory because even if the underlying distribution is not normal (i.e. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$; consistency property).

3) The spread in the distribution decreases with increasing sample size. This is as expected from theory because even if the underlying distribution is not normal (i.e. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.; spread).

4) The distributions appear to be shaped like a normal distribution when the underlying distribution is normal and uniform, and the skewness is

close to zero in all sample sizes. But for the exponential distribution the sample means have distributions that are more symmetric, as the sample size increases. The skewness is close to zero at n=100.

5) The fitting distribution test for normality “Kolmogorov-Smirnov d” is **not significant** in all sample sizes when the underlying distribution is normal and uniform. So we conclude that we can’t reject the null hypothesis that states “**the distribution is fitted to the normal distribution”.** That is the sampling distribution of the sample mean is normally distributed for the selected sample sizes.

But when the underlying distribution is exponential, this test of normality is significant for small sample sizes, but when n=100, we conclude that test is not significant (i.e. the sampling distribution of $\overline{X}$ is approximately normal).

6) The simulated results using both the Probabilistic simulation method and the Bootstrap method arrived at the same conclusions.

a) $\mu = \mu_{\bar{x}}$

b) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

c) The shape of $\overline{x}$ is approximately normal.

Table (3.1.1)

Probabilistic simulation results table for the sampling distribution of the sample mean: N (0, 1).

SAMPLE SIZE n	THEORITICAL (POPULATION) MEAN μ	SIMULATED MEAN OF MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE KOLMOGOROV-SMIRNOV NORMALITY TEST
n=2	0	-0.01	0.71	0.74	-.19	0.33	P= n.s
n=5	0	-0.01	0.45	0.46	-.14	-0.13	P= n.s
n=10	0	0.00	0.32	0.32	-.04	-0.17	P= n.s
n=15	0	0.00	0.26	0.26	0.03	-0.02	P= n.s
n=20	0	0.00	0.22	0.23	0.04	0.06	P= n.s
n=30	0	0.01	0.18	0.19	0.04	-0.04	P= n.s
n=50	0	0.00	0.14	0.15	0.04	0.32	P= n.s
n=100	0	0.00	0.10	0.10	0.01	0.00	P= n.s

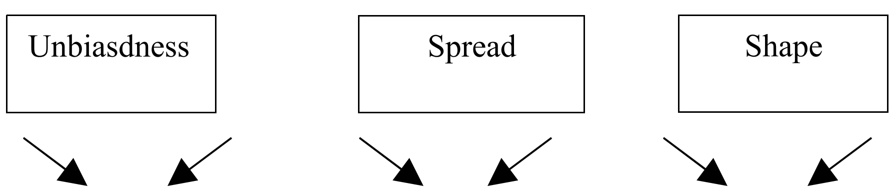


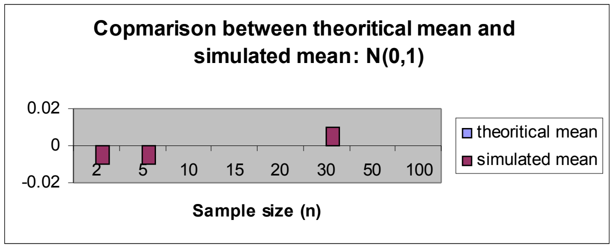
Table (3.1.2)

Bootstrap Simulation results for the sampling distribution of the sample mean: N (0, 1).

SAMPLE SIZE n	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN OF THE MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV SMIRNOV NORMALITY FITTING TEST
n=2	0	0.08	0.71	0.69	0.00	-0.57	P=n.s
n=5	0	0.04	0.45	0.45	-0.07	-0.26	P=n.s
n=10	0	0.02	0.32	0.31	-0.01	-0.01	P=n.s
n=15	0	0.02	0.23	0.25	-0.02	-0.01	P=n.s
n=20	0	0.02	0.22	0.21	0.03	-0.02	P=n.s
n=30	0	0.02	0.18	0.18	0.02	-0.14	P=n.s
n=50	0	0.02	0.14	0.14	0.00	0.07	P=n.s
n=100	0	0.02	0.10	0.09	0.00	0.04	P=n.s

Figure (1.1)

Sampling distribution of the sample mean: N (0, 1)



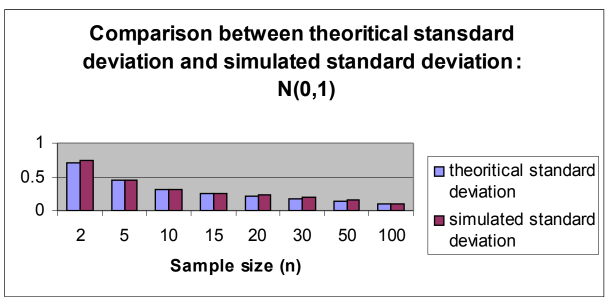


Table (3.1.3)

Bootstrap Simulation results for the sampling distribution 0f the mean: U (10, 20)

SAMPLE SIZE n	THEORITICAL (POPULATION) MEAN μ	SIMULATED MEAN OF THE MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV SMIRNOV NORMALITY FITTING TEST
n=2	15	15.05	2.04	2.04	0.102	-0.572	P=n.s
n=5	15	15.04	1.34	1.32	0.080	-0.178	P=n.s
n=10	15	15.08	0.93	0.93	0.099	-0.013	P=n.s
n=15	15	15.08	0.73	0.73	0.046	-0.145	P=n.s
n=20	15	15.08	0.65	0.65	-0.068	0.079	P=n.s
n=30	15	15.01	0.53	0.53	-0.070	0.057	P=n.s
n=50	15	15.10	0.42	0.42	-0.014	-0.090	P=n.s
n=100	15	15.10	0.29	0.29	-0.102	-0.294	P=n.s

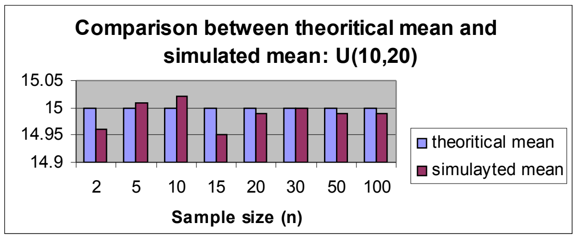
Table (3.1.4)

Probabilistic Simulation results for the sampling distribution of the mean: U (10, 20).

SAMPLE SIZE	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN OF THE MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD. DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV NORMALITY FITTING TEST
n=2	15	14.96	2.04	2.04	-0.11	-0.52	P=n.s
n=5	15	15.01	1.29	1.27	-0.06	-0.09	P=n.s
n=10	15	15.02	0.91	0.91	0.05	0.01	P=n.s
n=15	15	14.95	0.75	0.74	0.05	-0.03	P=n.s
n=20	15	14.99	0.65	0.65	0.09	-0.03	P=n.s
n=30	15	15.00	o.53	0.53	0.06	0.09	P=n.s
n=50	15	14.99	0.41	0.40	-0.05	-0.07	P=n.s
n=100	15	14.99	0.29	0.28	0.01	-0.01	P=n.s

Figure (1.2)

Sampling distribution of the sample mean: U (10, 20)



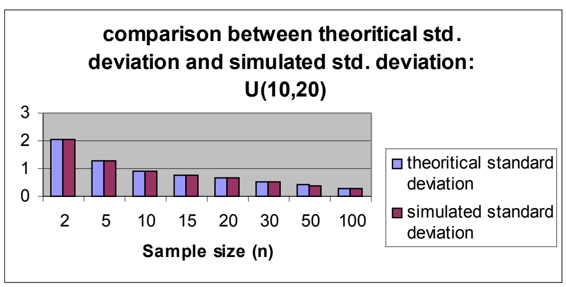


Table (3.1.5)

Bootstrap Simulation results for the sampling distribution of the mean: EXP (2).

SAMPLE SIZE n	THEORITICAL (POPULATION) MEAN μ	SIMULATED MEAN OF THE MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV SMIRNOV NORMALITY FITTING TEST
n=2	2	1.87	1.41	1.37	1.36	1.97	P<0.01
n=5	2	1.89	0.89	0.86	0.78	0.38	P<0.01
n=10	2	1.88	0.63	0.62	0.72	0.71	P<0.01
n=15	2	1.86	0.52	0.49	0.56	0.26	P<0.01
n=20	2	1.85	0.45	0.43	0.52	0.39	P<0.01
n=30	2	1.85	0.37	0.35	0.51	0.34	P<0.01
n=50	2	1.87	0.28	0.26	0.37	0.06	P<0.01
n=100	2	1.87	0.20	0.18	0.18	-0.05	P=n.s

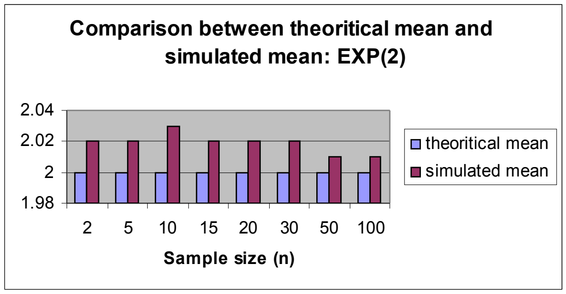
Table (3.1.6)

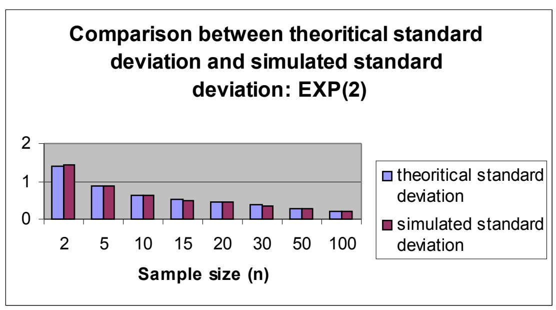
Probabilistic Simulation results for the sampling distribution of the mean: EXP (2).

SAMPLE SIZE n	THEORETICAL MEAN (POPULATION) μ	SIMULATED MEAN OF THE MEANS $\mu_{\bar{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\bar{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV SMIRNOV NORMALITY FITTING TEST
n=2	2	2.02	1.40	1.44	1.35	2.62	P<.01
n=5	2	2.02	0.89	0.88	0.79	1.02	P<.01
n=10	2	2.03	0.63	0.62	0.45	0.16	P<.01
n=15	2	2.02	0.52	0.50	0.34	-0.06	P<.01
n=20	2	2.02	0.45	0.44	0.26	-0.08	P<.01
n=30	2	2.02	0.37	0.36	0.25	0.36	P<.01
n=50	2	2.01	0.28	0.28	0.32	-0.11	P<.05
n=100	2	2.01	0.20	0.20	0.20	0.11	P=n.s

Figure (1.3)

Sampling distribution of the sample mean: EXP (2)





3.2 simulation procedure and sampling distribution of the sample median:

In this section we have studied graphically, computationally and analytically the sampling distribution of the sample median using the two methods of simulation, the Probabilistic and the Bootstrap methods. We have used three distributions N (0, 1), U (10, 20) and EXP (2).

Simulation procedure:

1000 samples were drawn from the above-mentioned distributions with sample sizes n=2, 5, 10, 15, 20, 30, 50 and 100. Then histograms were drawn to explore the sampling distribution of the sample median with parameters $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}$.

Kolmogorov-smirnov goodness of fit test for normality was used in each case to assess the fit.

Using simulations and in particular the histograms in appendix 2, the tables (3.2.1-3.2.6) and the bar charts (2.1-2.3), we arrive at the following conclusions:

1) The sample mean of the medians (simulated mean of medians) from normal, uniform and exponential distribution is approximately equal to the population mean regardless of the sample size. This agrees with the

theory that a sample median is an unbiased estimator of the population mean regardless of the underlying population distribution (i.e. $\mu = \mu_{\tilde{x}}$)

2) Simulated Standard deviation of the medians is approximately equal to the population standard deviation times $1.25 \frac{1}{\sqrt{n}}$ regardless to the sample sizes and when samples drawn from normal, uniform and exponential for all sample sizes. This is as expected from theory because even if the underlying distribution is not normal (i.e. $\sigma_{\tilde{x}} = 1.25 \frac{\sigma}{\sqrt{n}}$).

3) The spread in the distribution decreases with increasing sample size. This is as expected from theory because even if the underlying distribution is not normal

4) The distributions appear to be shaped like a normal distribution when the underlying distribution is normal and uniform, and the skewness is close to zero in all sample sizes. But for the exponential distribution the sample means have distributions that are more symmetric, as the sample size increases. The skewness is close to zero at n=100.

5) The fitting distribution test for normality "kolmogorov-smirnov d" is **not significant** in all sample sizes when the underlying distribution is

normal and uniform. So we conclude that we can't reject the null hypothesis that says "**the distribution is fitted to the normal distribution".** That is the sampling distribution of the sample medians is normally distributed for the selected sample sizes.

But when the underlying distribution is exponential, this test of normality is significant for all sample sizes but when n=100 and larger sizes, we conclude that test is not significant (i.e. the sampling distribution of the median is approximately normal)

6) The simulated results using both probabilistic simulation method and the bootstrap method arrived at the same conclusions.

1) $\mu = \mu_{\tilde{x}}$.

2) $\sigma_{\tilde{x}} = 1.25 \frac{\sigma}{\sqrt{n}}$.

3) The shape of $\tilde{x}$ is normal.

Table (3.2.1)

Probabilistic Simulation result for the sampling distribution of the sample median: N (0, 1)

SAMPLE SIZE	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE FOR KOLMOGOROV-SMIRNOV NORMALITY TEST
n=2	0	0.00	0.88	0.74	-0.19	0.33	P= n.s
n=5	0	-0.01	0.56	0.55	0.00	0.09	P= n.s
n=10	0	-0.01	0.39	0.38	0.00	0.04	P= n.s
n=15	0	0.00	0.32	0.32	0.08	0.07	P= n.s
n=20	0	0.00	0.30	0.32	0.08	0.04	P= n.s
n=30	0	0.00	0.23	0.27	0.07	0.05	P= n.s
n=50	0	0.00	0.18	0.18	0.06	0.34	P= n.s
n=100	0	0.00	0.13	0.13	0.02	0.21	P= n.s

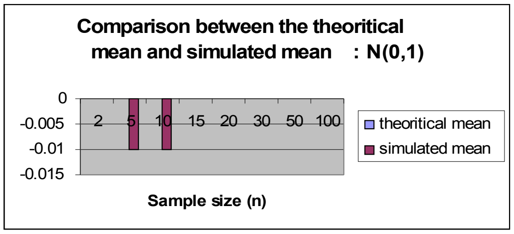
Table (3.2.2)

Bootstrap Simulation results for the sampling distribution of the sample median: N (0, 1).

SAMPLE SIZE	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEW NESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV NORMALITY FITTING TEST.
n=2	0	0.08	0.88	0.69	0.00	-0.56	P=n.s
n=5	0	0.06	0.56	0.55	-0.07	0.35	P=n.s
n=10	0	0.04	0.39	0.38	-0.06	1.09	P=n.s
n=15	0	0.04	0.32	0.32	-0.09	0.62	P=n.s
n=20	0	0.03	0.28	0.27	-0.08	0.39	P=n.s
n=30	0	0.04	0.23	0.23	-0.02	-0.39	P=n.s
n=50	0	0.04	0.18	0.18	0.08	-1.24	P=n.s
n=100	0	0.03	0.13	0.12	0.02	-1.87	P=n.s

Figure (2.1)

Sampling distribution of the sample median: N (0, 1)



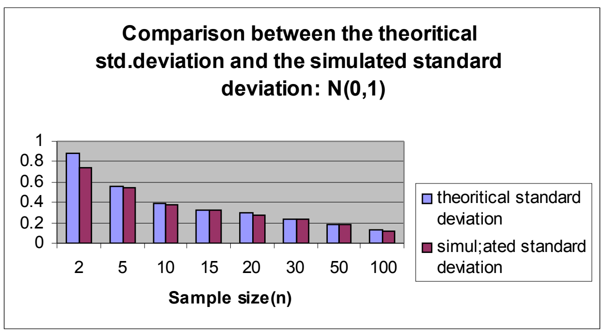


Table (3.2.3)

Probabilistic Simulation result for the sampling distribution of the sample median: U (10, 20).

SAMPLE SIZE	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORETICAL STD.DEVIATION	SIMULATED STD.DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV TEST FOR NORMALITY

n=2	15	14.96	2.50	2.04	-0.11	-0.53	P=n.s
n=5	15	15.02	1.65	1.90	-0.02	-0.65	P=n.s
n=10	15	14.99	1.15	1.40	-0.03	-0.24	P=n.s
n=15	15	14.97	0.95	1.20	0.13	-0.35	P=n.s
n=20	15	14.96	0.81	1.04	0.09	-0.28	P=n.s
n=30	15	14.97	0.65	0.88	0.04	-0.03	P=n.s
n=50	15	14.98	0.51	0.67	0.09	-0.12	P=n.s
n=100	15	14.97	0.37	0.47	0.09	0.22	P=n.s

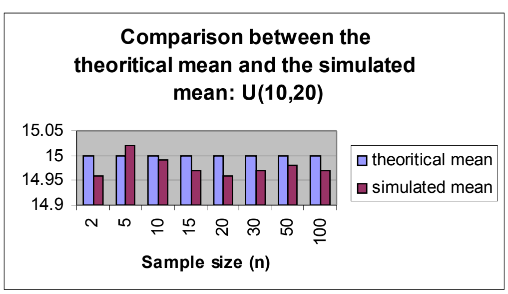
Table: (3.2.4)

Bootstrap Simulation results for the sampling distribution of the sample median: U (10, 20).

SAMPLE SIZE	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEW NESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV NORMALITY FITTING TEST .
n=2	15	15.05	2.04	2.55	0.102	-0.572	P=n.s
n=5	15	15.05	1.93	1.62	0.053	-0.756	P=n.s
n=10	15	15.10	1.43	1.15	-0.123	-0.605	P=n.s
n=15	15	15.12	1.27	0.94	-0.199	-0.585	P=n.s
n=20	15	15.15	1.09	0.81	-0.246	-0.543	P=n.s
n=30	15	15.20	0.93	0.66	-0.299	-0.388	P=n.s
n=50	15	15.22	0.75	0.52	-0.192	-0.385	P=n.s
n=100	15	15.26	0.55	0.36	0.260.263	-0.338	P=n.s

Figure (2.2)

Sampling distribution of the sample median: N (0, 1)



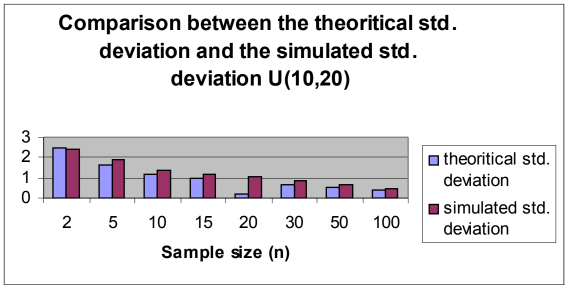


Table (3.2.5)

Probabilistic Simulation results for the sampling distribution of the sample median: EXP (2).

SAMPLE SIZE n	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORETICAL STANDARD DEVIATION	SIMULATED STANDARD DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEW NESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV NORMALITY FITTING TEST
n=2	2	2.02	1.76	1.44	1.35	2.80	P<0.01
n=5	2	1.55	1.12	0.93	1.28	2.75	P<0.01
n=10	2	1.49	0.79	0.62	0.78	0.88	P<0.01
n=15	2	1.47	0.65	0.55	0.76	0.52	P<0.01
n=20	2	1.46	0.56	0.45	0.61	0.62	P<0.01
n=30	2	1.44	0.46	0.37	0.60	0.71	P<0.01
n=50	2	1.42	0.35	0.29	054	0.55	P<0.01
n=100	2	1.40	0.25	0.20	0.43	0.51	P=n.s

Table: (3.2.6)

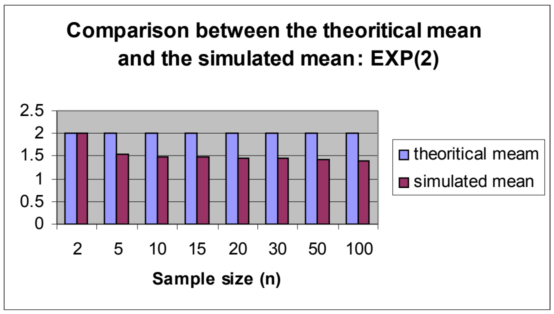
Bootstrap Simulation results for the sampling distribution of the sample median: EXP (2).

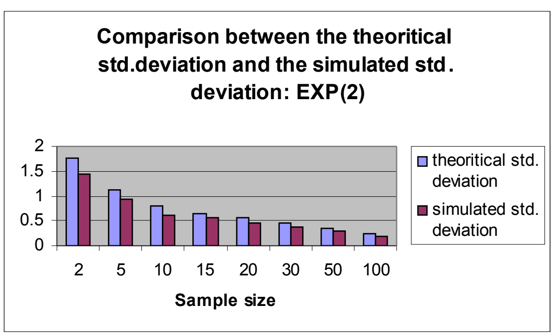
SAMPLE SIZE n	THEORETICAL (POPULATION) MEAN μ	SIMULATED MEAN $\mu_{\hat{x}}$	THEORITICAL STANDAD DEVIATIN	SIMULATD STANDARD DEVIATION $\sigma_{\hat{x}}$	SIMULATED SKEW NESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV NORMALITYFITTING TEST
n=2	2	1.87	1.77	1.37	1.361	1.970	P<0.01
n=5	2	1.47	1.12	0.89	1.163	1.348	P<0.01
n=10	2	1.39	0.7	0.58	1.003	1.780	P<0.01

n=15	2	1.34	0.65	0.49	0.893	1.209	P<0.01
n=20	2	1.33	0.56	0.41	0.923	1.877	P<0.01
n=30	2	1.30	0.46	0.33	0.634	0.759	P<0.01
n=50	2	1.31	0.35	0.25	0.572	0.936	P<0.01
n=100	2	1.30	0.25	0.17	0.217	0.179	P=n.s

Figure (2.3)

Sampling distribution of the sample median: EXP (2)





3.3 Simulation procedure and the sampling distribution of the sample variance $\frac{(n-1)s_i^2}{\sigma^2}$:

In this section we have studied analytically the sampling distribution of the quantity $\frac{(n-1)s_i^2}{\sigma^2}$ using two methods of simulation, probabilistic method and bootstrap method. We have used three distributions: N (0, 1), U (10, 20) and EXP (2).

Simulation procedure:

1000 samples were drawn from the above mentioned distributions with sample sizes n=2, 5, 10, 15, 20, 30, 50 and 100. Histograms were drawn to explore the sampling distribution of $\frac{(n-1)s_i^2}{\sigma^2}$. Kolmogorov-smirnov goodness of fit test for the chi-square with (n-1) degrees of freedom has been used in each case.

Using simulations and in particular the histograms in appendix 3 and the tables (3.3.1-3.3.6), we arrive at the following conclusions:

The shape of the sampling distribution of $\frac{(n-1)s_i^2}{\sigma^2}$ is positively skewed for n=2, 5, 10, 15, 20 when the underlying distribution is normal; since the skewness is far from zero which indicates that the population is not symmetric. For n=30, 50, 100 regardless of the underlying distribution, the shape of distribution is normal, because the skewness is close to zero in these cases. This is as expected from theory because with large samples, the sampling distribution is approximately normal even though the underlying distribution is uniform distribution or exponential.

2) The fitting distribution test for chi-square (n-1), "kolmogorov smirnov d "**is not significant** at n=2, 5, 10, 15& 20 and 100 when the underlying distribution is normal. Therefore, we conclude that we cannot reject the null hypothesis that says the "**distribution is fitted to the chi-square distribution with (n-1) degrees of freedom".**

This is as expected from theory because this quantity follows a chi-square distribution with n-1 degrees of freedom if the under lying distribution is normal.

Table (3.3.1)Probabilistic Simulation result for the sampling distribution of the sample variance: N (0, 1)

SAMPLE SIZE	DEGREES OF FREEDOM (N-1)	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOV-SMIRNOV CHI-SQUARE (N-1) TEST
n=2	1	2.82	12..1	n.s
n=5	4	1.36	2.77	n.s
n=10	9	0.99	1.56	n.s
n=15	14	0.79	0.99	n.s
n=20	19	0.59	0.77	n.s
n=30	29	0.43	0.31	0.01
n=50	49	0.46	0.35	0.01
n=100	99	0.19	-0.07	0.01

Table (3.3.2)Bootstrap Simulation results for the sampling distribution of the sample variance: N (0, 1)

SAMPLE SIZE	DEGREES OF FREEDOM (N-1)	SIMULATED SKEWNESS	KSIMULATED URTOSIS	P-VALUE OF THE KOLMOGOROV SMIRNOV CHI-SQUARE (N-1) FITTING TEST
n=2	1	2.74	9.86	n.s
n=5	4	1.22	1.84	n.s
n=10	9	1.12	0.83	n.s
n=15	14	0.99	0.29	n.s
n=20	19	0.89	0.09	n.s
n=30	29	0.59	-0.07	0.01
n=50	49	0.29	0.09	0.05
n=100	99	0.22	0.24	0.01

Table (3.3.3)

Probabilistic Simulation result of the sampling distribution of the sample variance: U (10, 20).

SAMPLE SIZE	DEGREES OF FREEDOM	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF KOLMOGOROV-SMIRNOV CHI-SQUARE (N-1)

				FITTING TESTS
n=2	1	1.47	1.60	0.01
n=5	4	0.38	-0.05	0.01
n=10	9	0.27	0.01	0.01
n=15	14	0.13	-0.19	0.01
n=20	19	0.06	-0.18	0.01
n=30	29	0.10	0.17	0.01
n=50	49	0.07	0.05	0.01
n=100	99	0.05	-0.07	0.01

Table (3.3.4)

Bootstrap Simulation results for the sampling distribution of the sample variance: U (10, 20).

SAMPLE SIZE	DEGREES OF FREEDOM (N-1)	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF THE KOLMOGOROV SMIRNOV CHI-SQUARE (N-1) FITTING TEST
n=2	1	1.633	2.290	0.01
n=5	4	0.415	-0.307	0.01
n=10	9	0.285	-0.164	0.01
n=15	14	0.189	-0.059	0.01
n=20	19	0.188	0.098	0.01
n=30	29	0.181	0.099	0.01
n=50	49	0.190	0.194	0.01
n=100	99	0.116	-0.201	0.01

Table (3.3.5)

Probabilistic Simulation results for the sampling distribution of the sample variance: EXP (2).

SAMPLE SIZE	DEGREES OF FREEDOM (N-1)	SIMULATED SKEWNESS	SIMULATED KURTOSIS	P-VALUE OF THE KOLMOGOROV SMIRNOV CHI-SQUARE (N-1) FITTING TEST
n=2	1	4.56	29.90	0.01
n=5	4	5.48	52.82	0.01
n=10	9	4.19	23.86	0.01
n=15	14	7.05	74.69	0.01
n=20	19	5.90	53.66	0.01
n=30	29	5.46	45.13	0.01
n=50	49	5.80	59.64	0.01
n=100	99	5.71	49.56	0.01

Table (3.3.6)

Bootstrapping Simulation results for sampling distribution of the sample variance: EXP (2)

SAMPLE SIZE	DEGREES OF	SIMULATED	SIMULATED	P-VALUE OF

	FREEDOM (N-1)	SKEW NESS	KURTOSIS	THE KOLMOGOROV SMIRNOV CHI-SQUARE (N-1) FITTING TEST
n=2	1	3.99	19.9	0.01
n=5	4	2.19	5.17	0.01
n=10	9	1.63	3.18	0.01
n=15	14	1.31	1.87	0.01
n=20	19	1.13	1.26	0.01
n=30	29	0.92	0.82	0.01
n=50	49	0.61	0.26	0.01
n=100	99	0.31	-0.06	0.01

3.4 Simulations of Independence of the sample mean and the sample variance:

In this section we have used the Probabilistic and Bootstrap simulation methods to explore the joint distribution of the simulated sample mean and the simulated sample variance based on an independent sample from normal (0, 1), uniform (10, 20) and exponential distribution (2). We have showed that the two statistics

$w_i = \frac{(n-1)s_i^{\,2}}{\sigma^2}$ and $v_i = \sqrt{n}\,\overline{x_i}$ are not correlated using the Pearson Correlation Coefficient.

Using simulations and in particular, looking at the tables (3.4.1-3.4.6) we arrived at the following conclusions:

w_i and v_i are not correlated when the underlying distribution is normal and uniform regardless of the sample size, since the Pearson correlation coefficient is not significant for each sample size.

However; Pearson correlation coefficients are significant when the underlying distribution is exponential. That means they are correlated

We notice the sample means and the sample variance are not correlated when the underlying distribution is symmetric for the cases N (0, 1) and U (10, 20).

Table (3.4.1)

Correlation between the sample mean and the sample variance:

N (0, 1)

Probabilistic Simulation Results			**Bootstrap Simulation Results**	
SAMPLE SIZE	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST
n=2	-.025	0.432	-0.058	0.07
n=5	0.001	0.567	-0.036	0.26
n=10	-0.017	0.600	-0.061	0.06
n=15	-0.120	0.710	-0.030	0.25
n=20	-0.017	0.590	-0.083	0.10
n=30	-0.017	0.590	-0.053	0.10
n=50	-0.001	0.970	-0.045	0.15
n=100	0.003	0.929	-0.050	0.11

Table (3.4.2)

Correlation between the sample mean and the sample variance:

U (10, 20)

Probabilistic Simulation Results			**Bootstrap Simulation Results**	
SAMPLE SIZE	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST
n=2	0.040	0.204	-0.021	0.503
n=5	0.025	0.427	-0.017	0.591
n=10	-0.081	0.100	-0.076	0.100
n=15	-0.019	0.549	-0.058	0.069
n=20	-0.006	0.858	-0.041	0.195
n=30	-0.004	0.906	-0.770	0.150
n=50	-0.011	0.739	-0.051	0.104
n=100	-0.012	0.694	-0.053	0.101

Table (3.4.3)

Correlation between the sample mean and the sample variance:

Exp (2)

Probabilistic Simulation Results			**Bootstrap Simulation Results**	
SAMPLE SIZE	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST	PEARSON CORRELATION COEFFICIENT BETWEEN W & V	P-VALUE OF THE CORRELATION TEST
n=2	0.627	.000	0.646	0.00
n=5	0.619	.000	0.726	0.00
n=10	0.601	.000	0.744	0.00
n=15	0.602	.000	0.738	0.00
n=20	0.633	.000	0.734	0.00
n=30	0.619	.000	0.736	0.00
n=50	0.601	.000	0.732	0.00
n=100	0.658	.000	0.719	0.00

SUMMARY OF THE SIMULATED RESULTS

Considering all the simulated results in this study, we conclude that, in large samples, the sampling distribution of the sample mean, the sample median and the sample variance can be approximated by a normal distribution regardless of the shape of the underlying distribution with a very small simulated bias.

In addition the theoretical standard deviation and the simulated standard deviation of these statistics get closer to each other as the sample size increases. However; the simulated standard deviation of the sample mean is smaller than the simulated standard deviation of the sample median.

For symmetric underlying distributions: normal and uniform distribution, the shape of the sample mean and sample median is approximately normal even for the small sample sizes ($n \leq 30$).

In this case, also the theoretical standard deviation and the simulated standard deviation of these statistics get closer to each other as the sample size increases.

For non symmetric underlying distributions, these approximations converge to normal for large samples. For non symmetric underlying distribution large samples are needed for convergence.

The simulated statistics of $\frac{(n-1)s_i^2}{\sigma^2}$ can be approximated by a chi-square distribution with n-1 degrees of freedom if the sample size is less than 30 and the underlying distribution is normal. Thus, symmetric underlying distribution is not adequate for this approximation.

Finally the simulated results shoed analytically that the sample mean and the sample variance are not correlated when the underlying distribution s are uniform or normal.

Chapter Four

Conclusion and Recommendations

4.1 Conclusion

In this thesis, we used the probabilistic and the bootstrap simulation methods to investigate the sampling distribution of the sample mean, the sample median, and the sample variance for different sample size and different underlying population distribution.

In particular, we used the above-mentioned methods of simulation to study graphically and computationally the two common properties of an estimator: bias and standard error. In addition, the shape of the sampling distribution of these statistics has been studied graphically and analytically using the histogram and the Kolmogorov-Smirnov goodness-of-fit test.

From the graphical, computational and the analytical simulated results using the probabilistic and the bootstrap methods, we conclude that in large samples, the sampling distribution of the sample mean, the sample median and the sample variance can be approximated by a normal distribution regardless of the shape of the underlying distribution. In addition, we can conclude that the simulated bias is very small, that for practical purpose and due to rounding errors can be considered equal to the population mean.

Upon comparing between the simulated standard deviation and the theoretical standard deviation of these statistics, we can conclude that these values get closer to each other as the sample size increases.

In addition, the simulated standard deviation of the sample mean is smaller than the simulated standard deviation of the sample median. This agrees with the theory that the sample mean has smaller variance than the sample median.

For the symmetric underlying distributions, normal and uniform, we can conclude graphically and analytically that the shape of the sampling distribution of the sample mean and the sample median is approximately normally distributed even for small sample size.

In addition, we can conclude that the simulated bias is very small, that for practical purpose and due to rounding errors can be considered equal to the population mean. Upon comparing between the simulated standard deviation and the theoretical standard deviation of these statistics, we can conclude that these values get closer to each other as the sample size increases.

In addition, the simulated standard deviation of the sample mean is smaller than the simulated standard deviation of the sample median. This agrees with the theory that the sample mean has smaller variance than the sample median.

But if the underlying distribution is not symmetric (i.e. the exponential distribution), we can conclude graphically and analytically that the shape of the sampling distribution of the sample mean and the sample median is approximately normally distributed for large sample size.

The graphical and the analytical simulated results indicated that the shape of the sampling distribution of the statistics $\frac{(n-1)s_i^2}{\sigma^2}$ is approximately chi-square with n-1 degrees of freedom for small sample size $(n \leq 30)$ and the underlying population is normal.

The simulated results showed that the sample mean and the sample variance are not correlated when the underlying population distributions are uniform or normal. However, the simulated results showed that these statistics are correlated when the underlying distribution is not symmetric.

4.2 Recommendations

Several questions have not been addressed in this thesis.

1) It should be of some interest to investigate the above results using several different types of resampling: permutation, cross-validation, and jackknife.
2) To investigate the above results using the bootstrap resampling method by considering other symmetric distribution to include the triangular distribution, the t distribution and other non-symmetric distributions to include the chi-square distribution, and the beta distribution.
3) To investigate the sampling distribution of other sample statistics to include, the correlation coefficient, and the slope of the simple linear regression.

Bibliography

1) Allan j. Nash (21/2/2001). Intermediate statistic lab retrieved 19/02/2004

From web site

http://www.spy.fua.edu/chez/ajn/class/sta31631/sec7.html

2) Arnholt, A. T. (1997), "Using Simulation as a Teaching Technique in Determining Power and Efficiency of Various Statistics," American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 143-147.*

3) Characteristics of sample statistics (n.d).retrieved 27/01/2004 from web site

http://www.tulane.edu/~irschick/lec_2_2004.pdf

4) Dambolena, I. G. (1986), "Using Simulation in Statistics Courses," *Collegiate Microcomputer*.

5) Daniel mc Fadden & Walter backer & angelica Eymann. (August 27, 2001). Statistical simulation.retreived 10/01/2004 from web site http://www.econ.bbk.ac.uk/faculty/beckert/statism7.pdf

6) David W.Stockburger (n.d). The sampling distribution. Retrieved 27/01/2004 from web site http://www.psychstat.smsu.edu/introbook/sbk19.htm.

7) Derchieh Hug (14 may 2000). Bootstrapping. News and views retrieved on 5/02/2004 from web site http://www.scc.ms.unimelb.edu.au/news/n14.html

8) Eric w. Weinstein. (n.d). Sample variance distribution. Retrieved 28/01/2004 from web site http://mathworld.wolfram.com/samplevariancedistribution.html

9) Joliffe, I.T. (1995). Sample sizes and the central limit theorem: the Poisson distribution as an illustration. The American statistician, 49, 269

10) Jamie D.Mills. (2002). Using computer simulation methods to teach statistics. Journal of statistics education. Retrieved 10/6/2003 from web site

http://www.amstat.org/pupication/jse/v10n1/mills.html

11) Kersten, T. (1983), "Computer Simulations to Clarify Key Ideas of Statistics," *Two-Year College Mathematics Journal*.

12) M. Davidian. (Spring 2004). Simulation studies in statistics. Retrieved 2/03/2004 from web site

http://www4.stat.ncsu.edu/~davedian/st810a/simulation_handout.pdf

13) Mendenhall W. and Sincich T. (1989). *Statistics for the Engineering and Computer* Sciences (2nd Ed.).

14) Ng, V. M., and Wong, K. Y. (1999), "Using Simulation on the Internet to Teach Statistics," *The Mathematics Teacher*.

15) Point estimator and interval estimation (n.d). Retrieved 20/5/2004 from web site

www.uci.edu/classes/pols/pols401/handouts/notes5.pdf.

16) Properties of good estimators (n.d). Retrieved 27/6/2004 from web site

http://ocw.mit.edu/NR/rdonlyres/civil-and-enviromental-engineering/1-017computiong .

17) Properties of normal sample (n.d). Retrieved 22/2/2004 from web site http://www.ds.unifi.it/VL/VL_EN/sample/sample6.html.

18) The bootstrap. (n.d). retrieved 5/12/2004 from web site http://www.matchcs.duq.edu./larget/math496/bootstrap.

19) Tim c. Hester burg (n.d). Simulation and bootstrapping for teaching statististics. Retrieved 12/3/2004 from web site http://www.statsci.com/hesterburg

20) Sampling distribution and confidence interval (n.d) retrieved 22/9/2003 http://www.cis.ysu.edu/~chang/class/LN3743_5.PDF

21) Schwarz, C. J., and Sutherland, J. (1997), "Journal *of Statistics Education*. (www.amstat.org/publications/jse/v5n1/schwarz.html) .

22) Simulation to illustrate independents of sample mean and variance for

normal data (n.d). Retrieved 20/4/2004 from web site

http://mathworld.wolfram.com/samplevariancedistribution.html

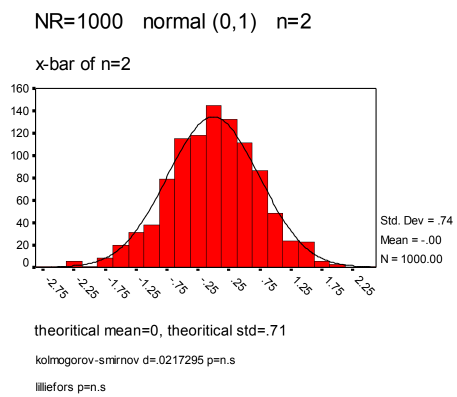
23) Weir, C. G., McManus, I. C., and Kiely, B. (1990), "Evaluation of the Teaching of Statistical Concepts by Interactive Experience with Monte Carlo Simulations," *British Journal of Educational Psychology*

24) West, R. W., and Ogden, R. T. (1998), "Interactive Demonstrations for Statistics Education on the World Wide Web," Journal of Statistics Education. (www.amstat.org/publications/jse/v6n3/west.html).

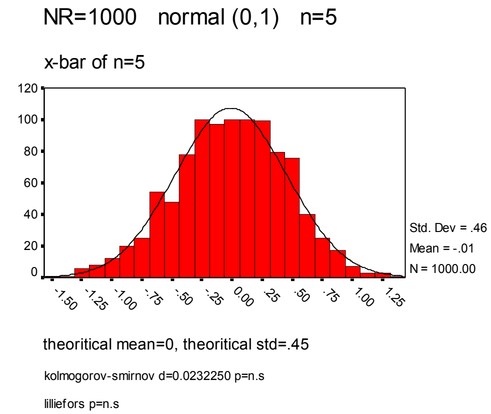
25) Yu, C., Behrens, J. T., and Anthony, S. (1995), "Identification of Misconceptions in the Central Limit Theorem and Related Concepts and Evaluation of Computer Media as a Remedial Tool,", New Orleans, LA.

Appendix 1

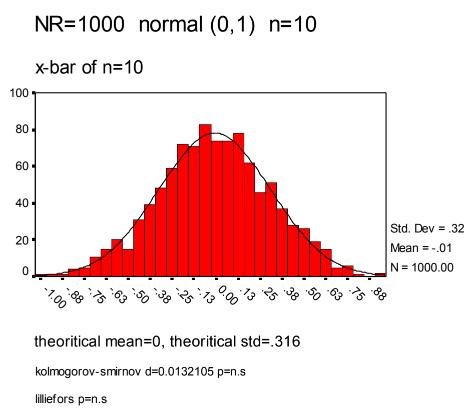
Figure (3.1.1) Probabilistic simulation histograms for the sampling distribution of the sample mean, N (0, 1)



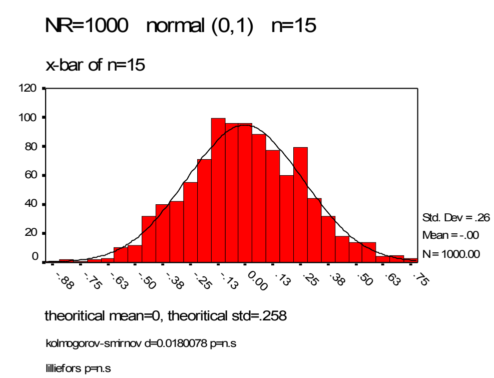
Histogram 1



Histogram 2



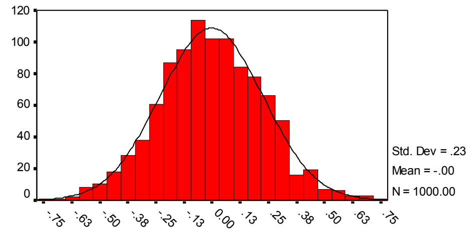
Histogram 3



Histogram 4

NR=1000 normal (0,1) n=20

x-bar of n=20



Histogram 5

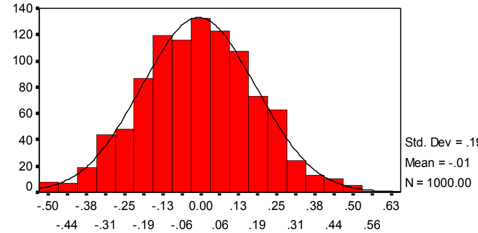
theoritical mean=0, theoritical std=.22

kolmogorov-smirnov d=0.0147266 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=30

x-bar of n=30



Histogram 6

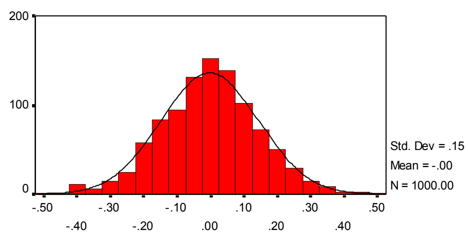
theoritical mean=0, theoritical std=.183

kolmogorov-smirnov d=0.0154383 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=50

x-bar of n=50



Histogram 7

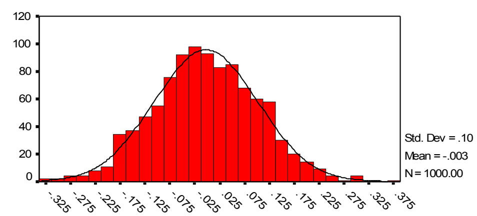
theoritical mean=0, theoritical std=.14

kolmogorov-smirnov d=0.0175863 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=100

x-bar of n=100



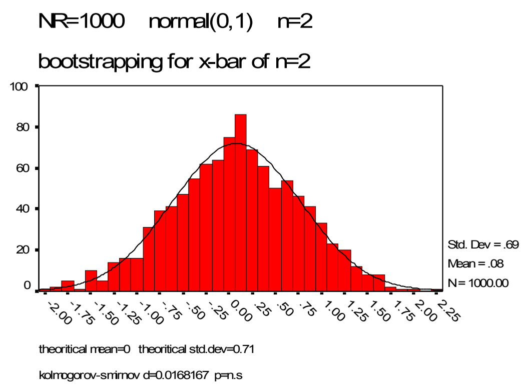
Histogram 8

theoritical mean=0, theoritical std=.1

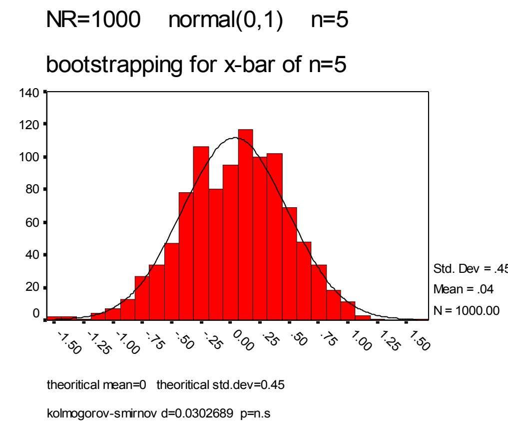
kolmogorov-smirnov d=0.0133079 p=n.s

lilliefors p=n.s

Figure (3.1.2) Bootstrap simulation histograms for the sampling distribution of the sample mean, N (0.1).

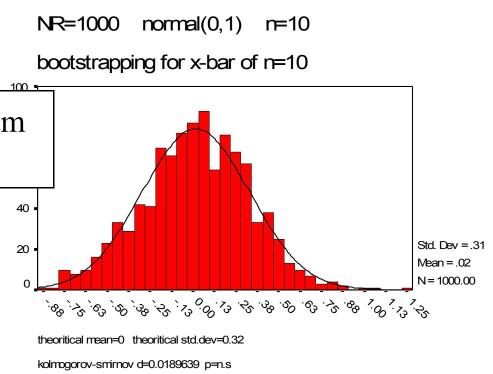


Histogram 9

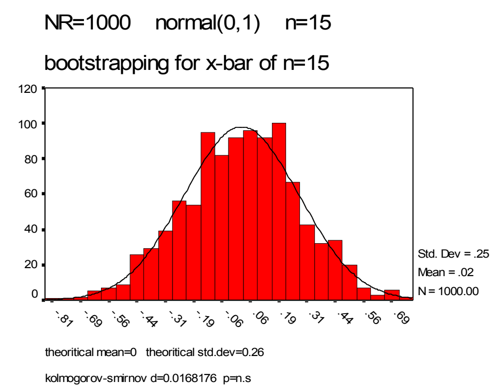


Histogram 10

Histogram 11

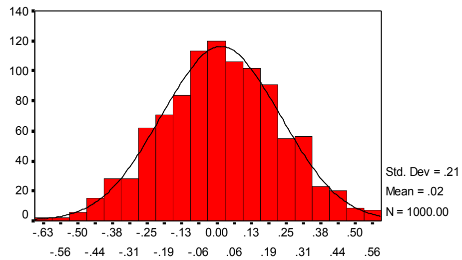


Histogram 12



NR=1000 normal(0,1) n=20

bootstrapping for x-bar of n=20



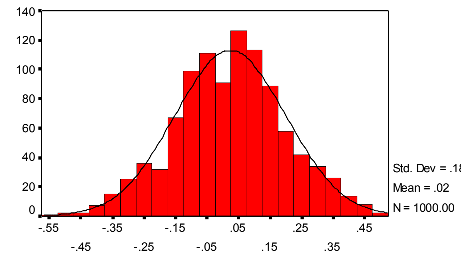
theoritical mean=0 theoritical std.dev=0.22

kolmogorov-smirnov d=0.0142481 p=n.s

Histogram 13

NR=1000 normal(0,1) n=30

bootstrapping for x-bar of n=30



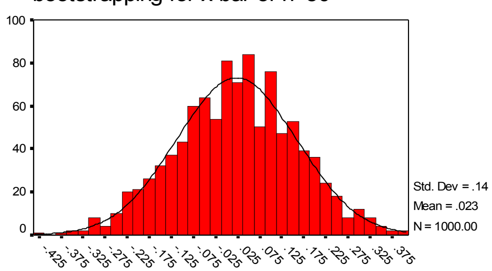
theoritical mean=0 theoritical std.dev=0.18

kolmogorov-smirnov d=0.0199631 p=n.s

Histogram 14

NR=1000 normal(0,1) n=50

bootstrapping for x-bar of n=50



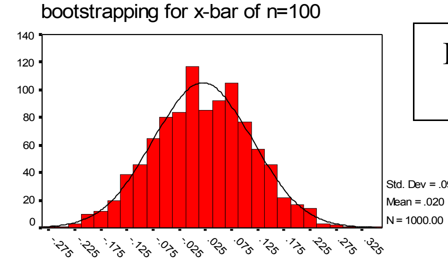
theoritical mean=0 theoritical std.dev=0.14

kolmogorov-smirnov d=0.0163755 p=n.s

Histogram 15

NR=1000 normal(0,1) n=100

bootstrapping for x-bar of n=100



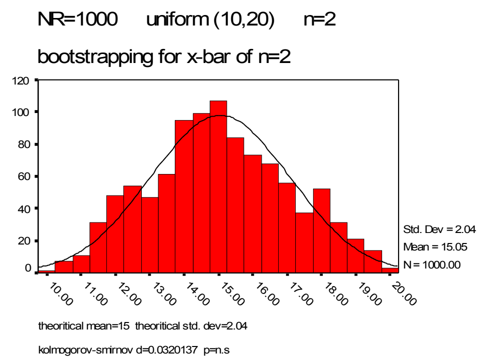
theoritical mean=0 theoritical std.dev=0.10

kolmogorov-smirnov d=0.0238846 p=n.s

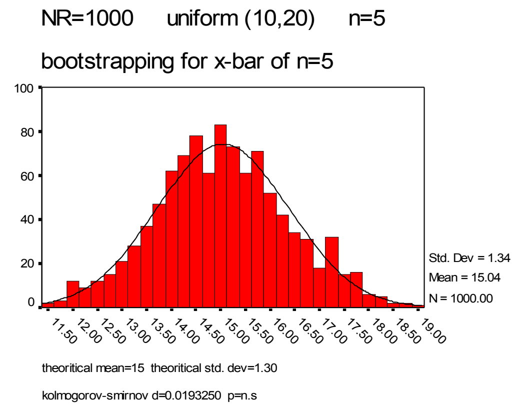
Histogram 16

Histogram 80

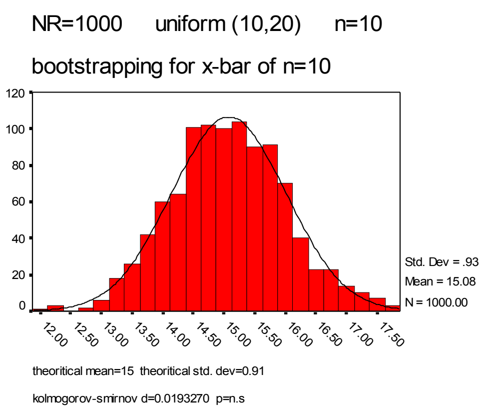
Figure (3.1.3) bootstrap simulation histograms for the sampling distribution of sample the mean, U (10, 20).



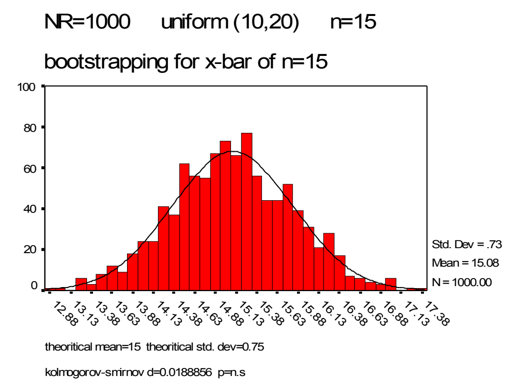
Histogram 17



Histogram 18



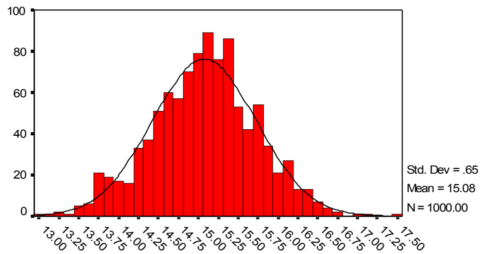
Histogram 19



Histogram 20

NR=1000 uniform (10,20) n=20

bootstrapping for x-bar of n=20



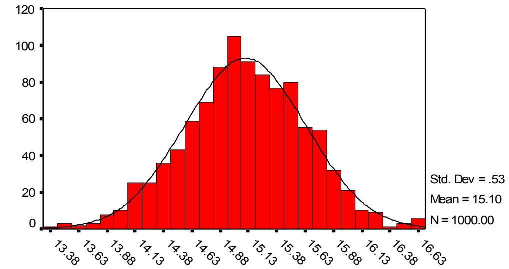
theoritical mean=15 theoritical std. dev=0.65

kolmogorov-smirnov d=0.232523 p=n.s

Histogram 21

NR=1000 uniform (10,20) n=30

bootstrapping for x-bar of n=30



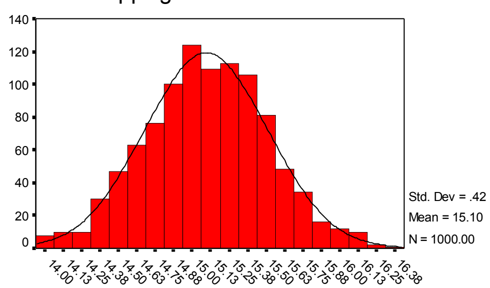
theoritical mean=15 theoritical std. dev=0.53

kolmogorov-smirnov d=0.0165218 p=n.s

Histogram 22

NR=1000 uniform (10,20) n=50

bootstrapping for x-bar of n=50



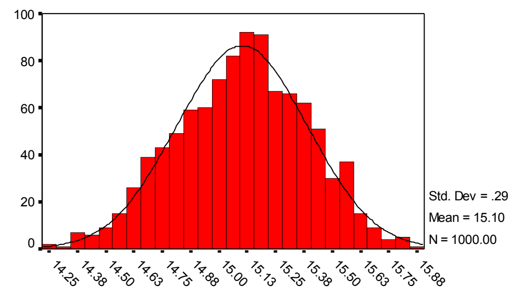
theoritical mean=15 theoritical std. dev=0.41

kolmogorov-smirnov d=0.0167286 p=n.s

Histogram 23

NR=1000 uniform (10,20) n=100

bootstrapping for x-bar of n=100

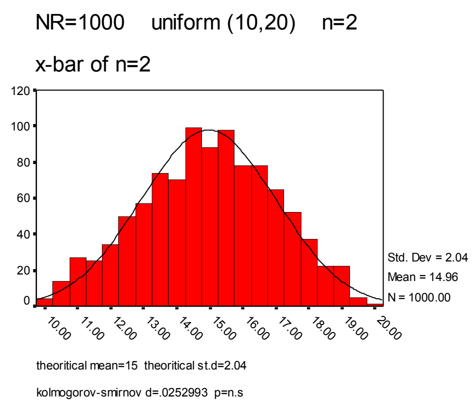


theoritical mean=15 theoritical std. dev=0.29

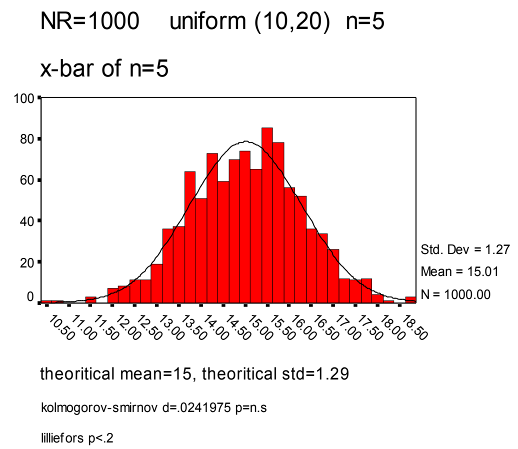
kolmogorov-smirnov d=0.067890 p=n.s

Histogram 24

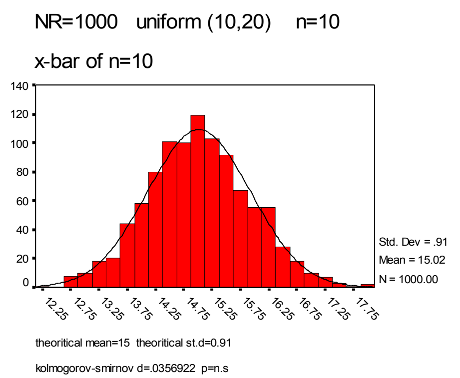
Figure (3.1.4) Probabilistic simulation histograms for the sampling distribution of the sample mean, U (10, 20)

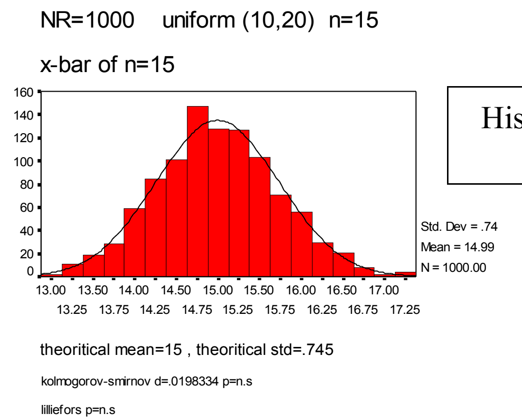


Histogram 25



Histogram 26



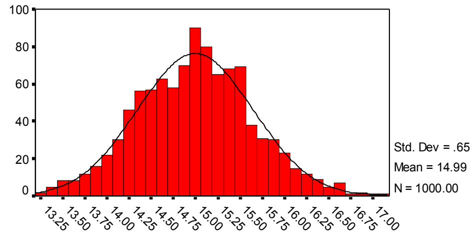


Histogram 27

Histogram 28

NR=1000 uniform (10,20) n=20

x-bar of n=20



Histogram 29

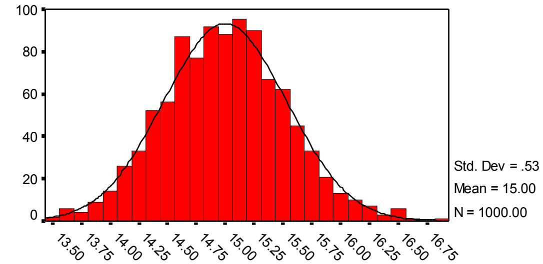
theoritical mean=15, theoritical std=.645

kolmogorov-smirnov d=.0.0179243 p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=30

x-bar of n=30



Histogram 30

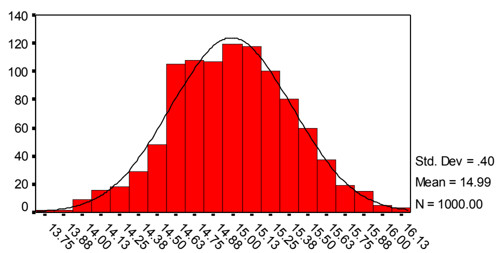
theoritical mean=15, theoritical std=.527

kolmogorov-smirnov d=.0181424 p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=50

x-bar of n=50



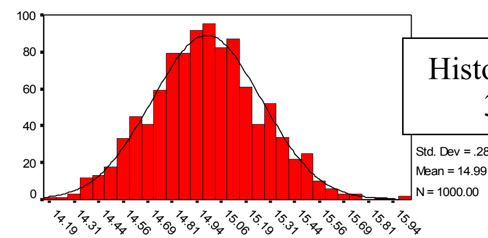
theoritical mean=15, theoritical std= .408

kolmogorov-smirnov d=.0.0183661 p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=100

x-bar of n=100



Histogram 31

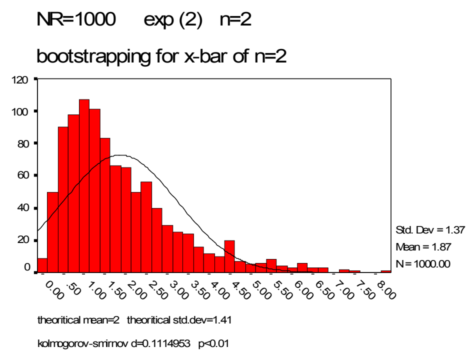
Histogram 32

theoritical mean=15, theoritical std=.288

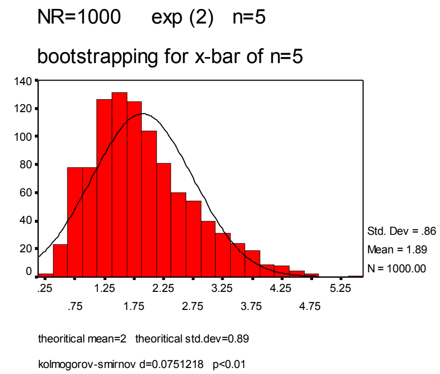
kolmogorov-smirnov d=.0.0172803 p=n.s

lilliefors p=n.s

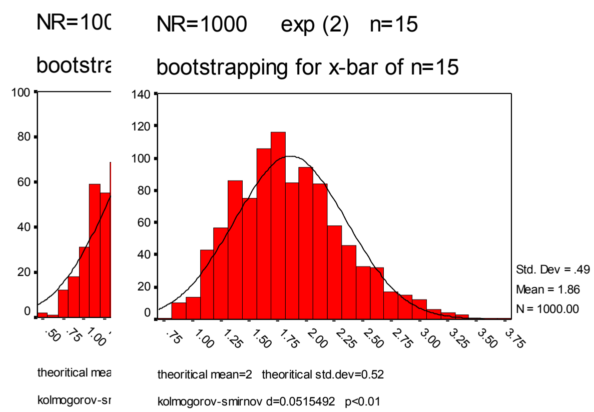
Figure (3.1.5) Bootstrap simulation histograms for the Sampling distribution of the sample mean, EXP (2).



Histogram 33



Histogram 34

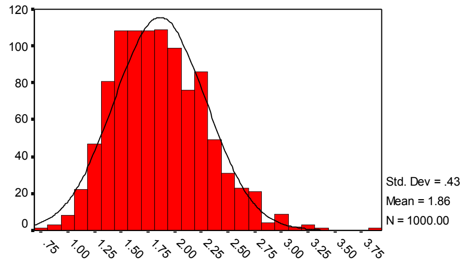


Histogram 35

Histogram 36

NR=1000 exp (2) n=20

bootstrapping for x-bar of n=20



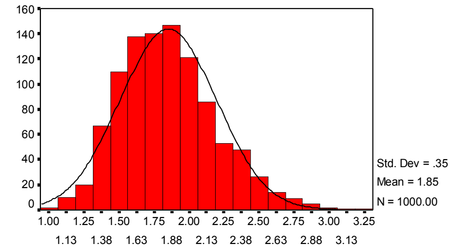
theoritical mean=2 theoritical std.dev=0.45

kolmogorov-smirnov d=0.0437895 p<0.01

Histogram 37

NR=1000 exp (2) n=30

bootstrapping for x-bar of n=30



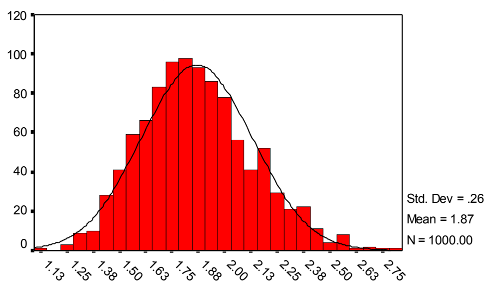
theoritical mean=2 theoritical std.dev=0.37

kolmogorov-smirnov d=0.0385356 p<0.01

Histogram 38

NR=1000 exp (2) n=50

bootstrapping for x-bar of n=50



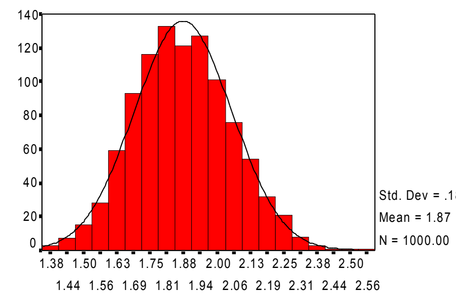
theoritical mean=2 theoritical std.dev=0.28

kolmogorov-smirnov d=0.0377072 p<0.01

Histogram 39

NR=1000 exp (2) n=100

bootstrapping for x-bar of n=100

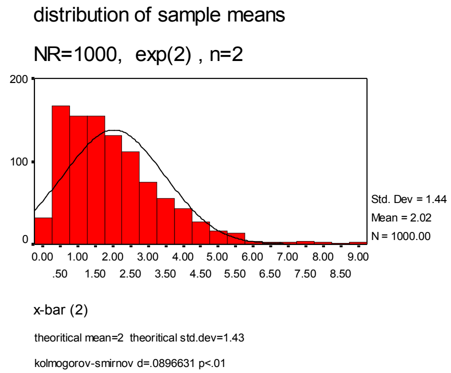


theoritical mean=2 theoritical std.dev=0.20

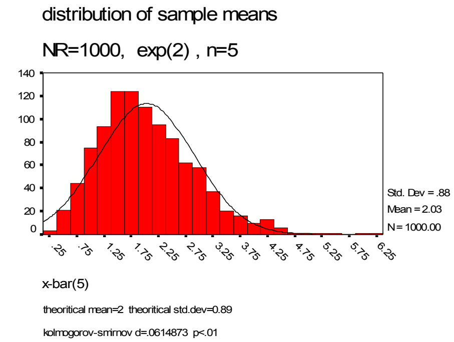
kolmogorov-smirnov d=0.0279407 p=n.s

Histogram 40

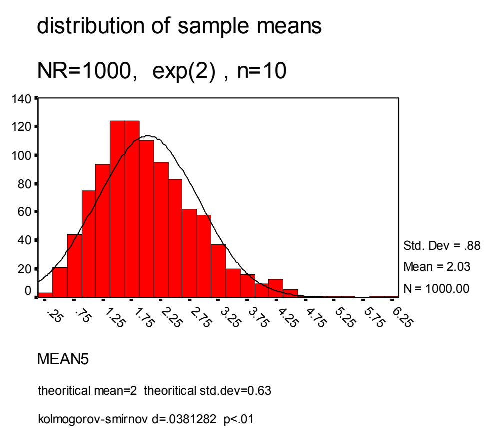
Figure (3.1.6) probabilistic simulation histograms for the sampling distribution of the sample mean, EXP (2).



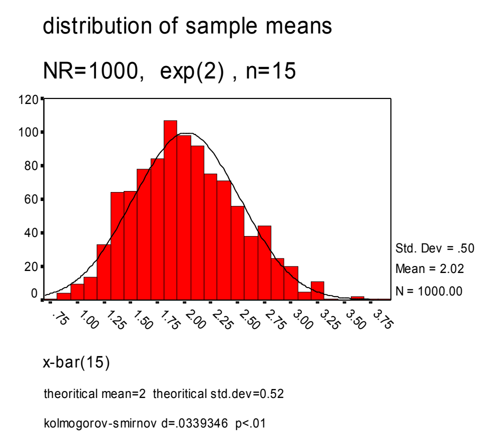
Histogram 41



Histogram 42



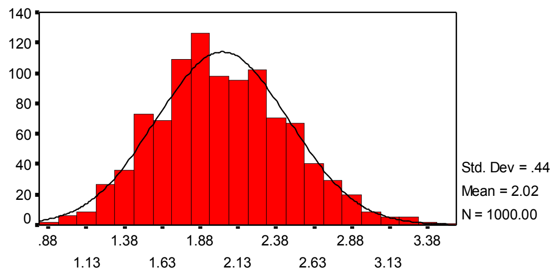
Histogram 43



Histogram 44

distribution of sample means

NR=1000, exp(2) , n=20



x-bar(20)

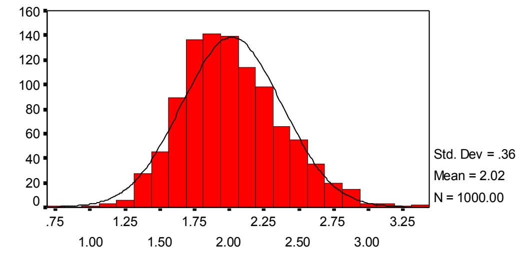
theoritical mean=2 theoritical std.dev=0.45

kolmogorov-smirnov d=.0387183 p<.01

Histogram 45

distribution of sample means

NR=1000, exp(2) , n=30



x-bar(30)

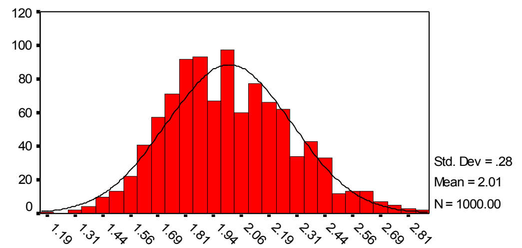
theoritical mean=2 theoritical std.dev=0.37

kolmogorov-smirnov d=.0463614 p<.01

Histogram 46

distribution of sample means

NR=1000, exp(2) , n=50



x-bar(50)

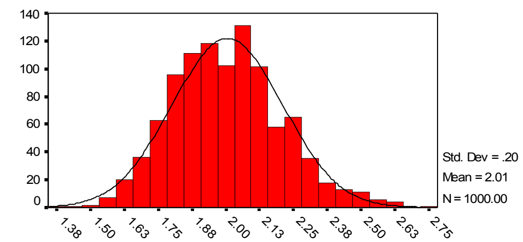
theoritical mean=2 theoritical std.dev=0.28

kolmogorov-smirnov d=.0463610 p<.05

Histogram 47

distribution of sample means

NR=1000, exp(2) , n=100



x-bar(100)

theoritical mean=2 theoritical std.dev=0.2

kolmogorov-smirnov d=.0273678 p=n.s

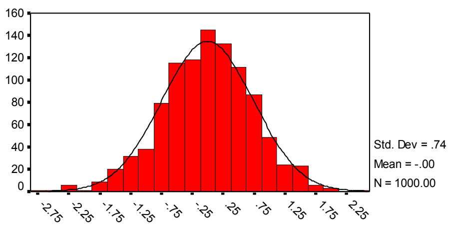
Histogram 48

Appendix 2

Figure (3.2.1) Probabilistic simulation histograms for Sampling distribution for the sample median, N (0, 1).

NR=1000 normal (0,1) n=2

median of n=2



Histogram 1

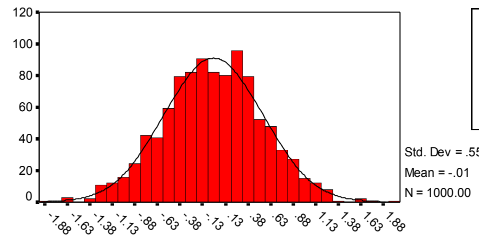
theoritical mean=0, theoritical std=.88

kolmogorov-smirnov d=0.0217295 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=5

median of n=5



Histogram 2

theoritical mean=0, theoritical std=.56

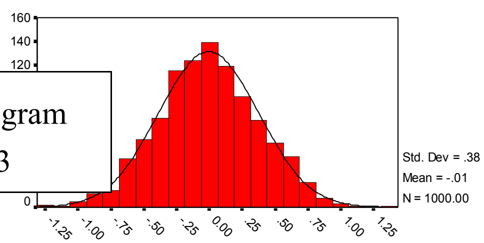
kolmogorov-smirnov d=0.0190525 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=10

median of n=10

Histogram 3



Histogram 4

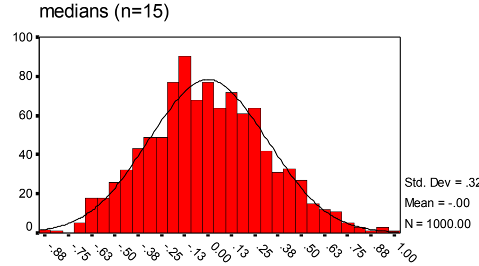
theoritical mean=0, theoritical std=.39

kolmogorov-smirnov d=0.0177135 p=n.s

lilliefors p=n.s

NR=1000 normal(1,0) n=15

medians (n=15)



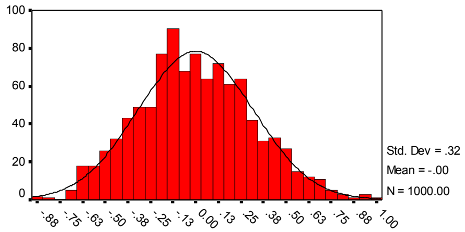
theoritical mean =0, theoritical std=.32

kolmogorov-smirnov d=.028813 p=n.s

lilliefors p= n.s

NR=1000 normal (0,1) n=20

median of n=20



Histogram 5

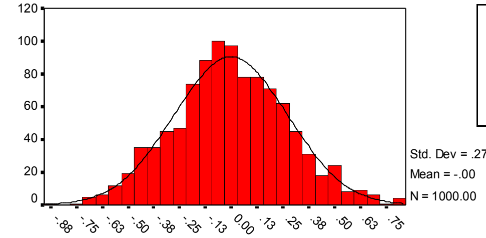
theoritical mean=0, theoritical std=.3

kolmogorov-smirnov d=0.0213664 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=30

median of n=30



Histogram 6

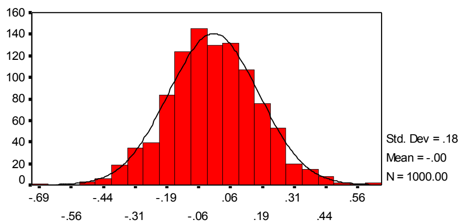
theoritical mean=0, theoritical std= .23

kolmogorov-smirnov d=0.0306508 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=50

median of n=50



Histogram 7

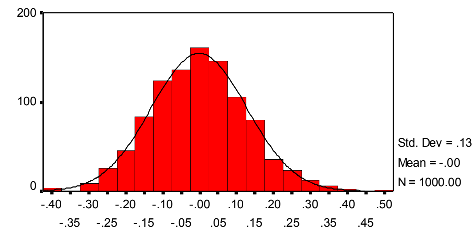
theoritical mean=0, theoritical std=.18

kolmogorov-smirnov d=0.0220220 p=n.s

lilliefors p=n.s

NR=1000 normal (0,1) n=100

median of n=100



Histogram 8

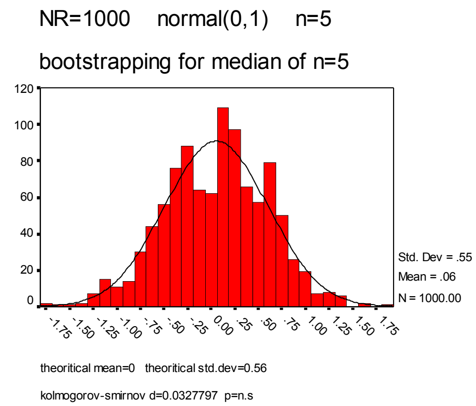
theoritical mean=0, theoritical std=.13

kolmogorov-smirnov d=0.0174597p=n.s

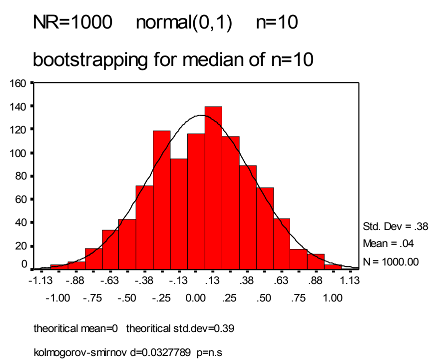
lilliefors p=n.s

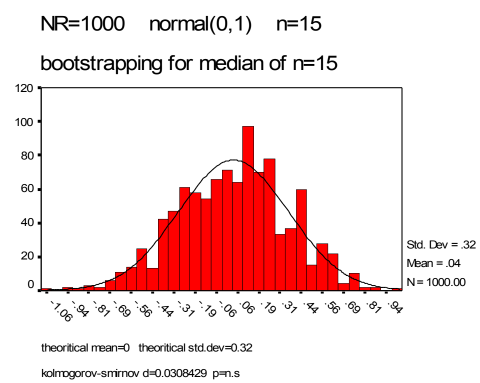
Figure (3.2.2) Bootstrap simulation histograms for Sampling distribution of the sample medians, N (0, 1). medians

Histogram
9



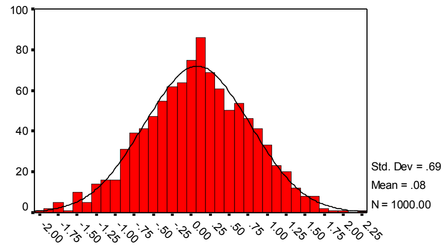
Histogram
10





NR=1000 normal(0,1) n=2

bootstrapping for median of n=2

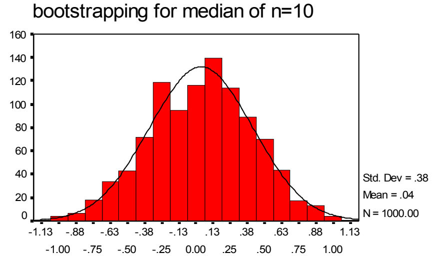


theoritical mean=0 theoritical std.dev=0.88

kolmogorov-smirnov d=0.0168167 p=n.s

NR=1000 normal(0,1) n=10

bootstrapping for median of n=10



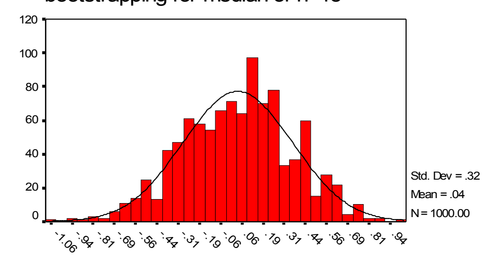
Histogram 11

theoritical mean=0 theoritical std.dev=0.39

kolmogorov-smirnov d=0.0327789 p=n.s

NR=1000 normal(0,1) n=15

bootstrapping for median of n=15



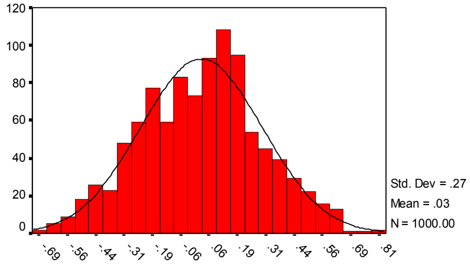
Histogram 12

theoritical mean=0 theoritical std.dev=0.32

kolmogorov-smirnov d=0.0308429 p=n.s

NR=1000 normal(0,1) n=20

bootstrapping for median of n=20



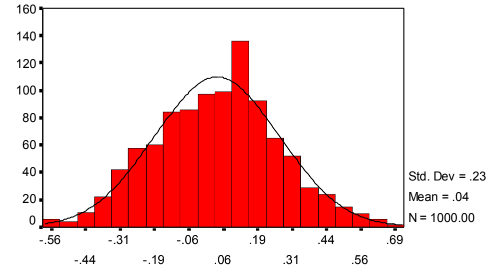
theoritical mean=0 theoritical std.dev=0.28

kolmogorov-smirnov d=0.0293469 p=n.s

Histogram 13

NR=1000 normal(0,1) n=30

bootstrapping for median of n=30



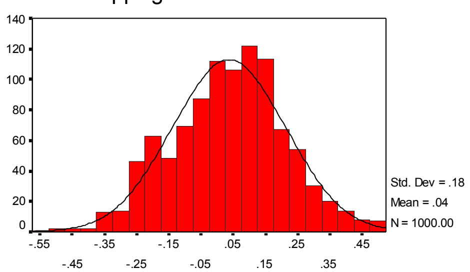
theoritical mean=0 theoritical std.dev=0.23

kolmogorov-smirnov d=0.0287697 p=n.s

Histogram 14

NR=1000 normal(0,1) n=50

bootstrapping for median of n=50



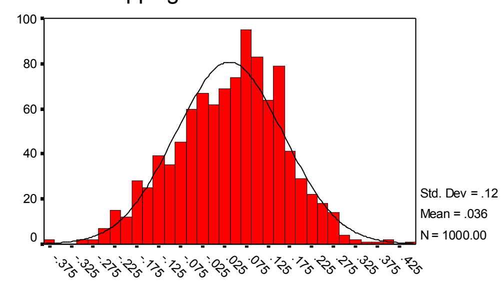
theoritical mean=0 theoritical std.dev=0.18

kolmogorov-smirnov d=0.0309965 p=n.s

Histogram 15

NR=1000 normal(0,1) n=100

bootstrapping for median of n=100



theoritical mean=0 theoritical std.dev=0.13

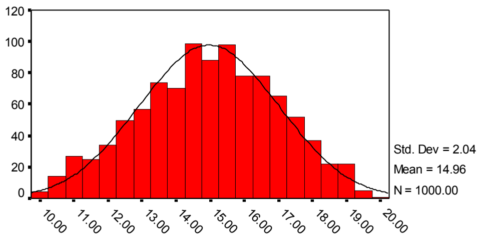
kolmogorov-smirnov d=0.0393436 p=n.s

Histogram 16

Figure (3.2.3) Probabilistic simulation histograms for the sampling distribution of the sample median: U (10, 20).

NR=1000 uniform (10,20) n=2

median of n=2



Histogram 17

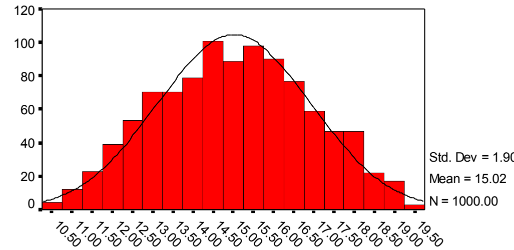
theoritical mean=15, theoritical std = 2.5

kolmogorov-smirnov d=.0252993, p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=5

median of n=5



Histogram 18

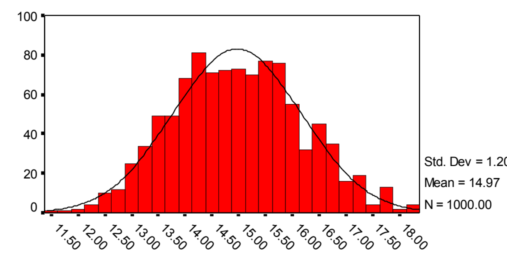
theoritical mean=15 , theoritical std= 1.65

kolmogorov-smirnov d=.0339502, p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=15

median of n=15



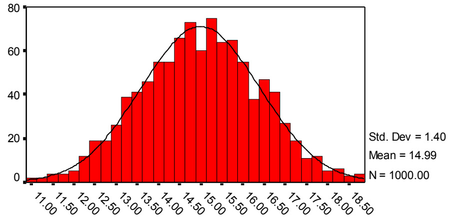
theoritical mean=15, theoritical std=.95

kolmogorov-smirnov d=.0291945 , p=n.s

lilliefors p=n.s

NR=1000 uniform (10,20) n=10

median of n=10



theoritical mean=15, theoritical std=1.15

kolmogorov-smirnov d=.0162491, p=n.s

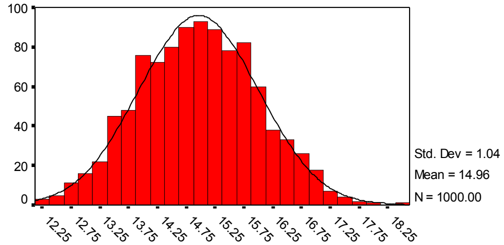
lilliefors p=n.s

Histogram 19

Histogram 20

NR=1000 uniform(10,20) n=20

median of n=20



theoritical mean=15, theoritical std=.807

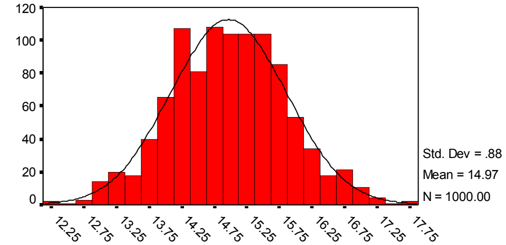
kolmogorov-smirnov d=.0213026 , p=n.s

lilliefors p=n.s

Histogram 21

NR=1000 uniform (10,20) n=30

median of n=30



theoritical mean=15, theoritical std=.65

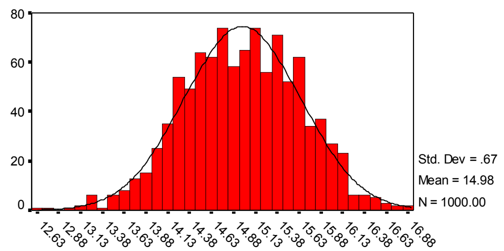
kolmogorov-smirnov d=.02227027 , p=n.s

lilliefors p=n.s

Histogram 22

NR=1000 uniform (10,20) n=50

median of n=50



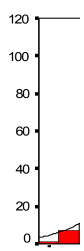
theoritical mean= 15, theoritical std=.51

kolmogorov-smirnov d=.022703 , p=n.s

lilliefors p=n.s

NR=

boot

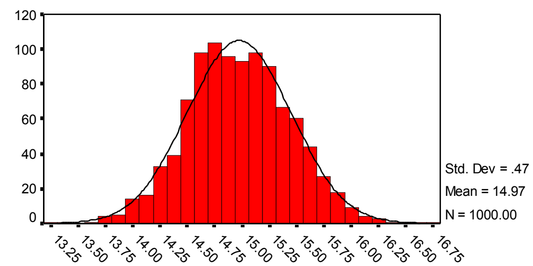


theoritical mean=15 theoritical std. dev=2.55

kolmogorov-smirnov d=0.0320137 p=n.s

NR=1000 uniform(10,20) n=100

median of n=100



theoritical mean=15, theoritical std=.37

kolmogorov-smirnov d=.0236835 , p=n.s

lilliefors p=n.s

theoritical mean=15 theoritical std. dev=1.62

kolmogorov-smirnov d=0.051715 p=n.s

Histogram 23

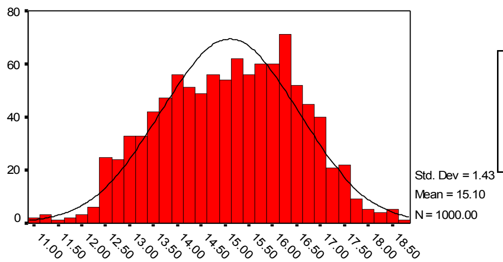
Histogram 24

Histogram 25

Histogram 26

NR=1000 uniform (10,20) n=10

bootstrapping for median of n=10



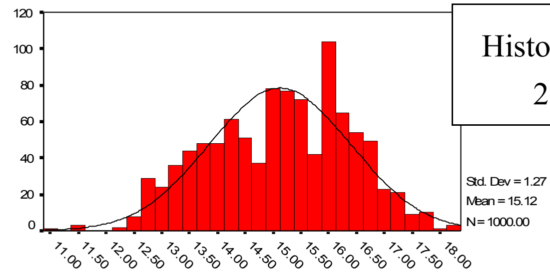
theoritical mean=15 theoritical std. dev=1.15

kolmogorov-smirnov d=0.0458539 p=n.s

Histogram 27

NR=1000 uniform (10,20) n=15

bootstrapping for median of n=15



Histogram 28

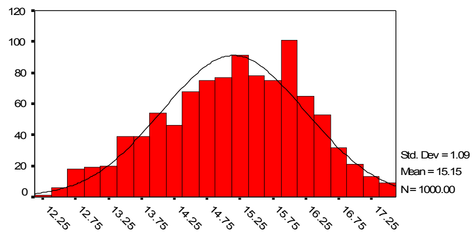
theoritical mean=15 theoritical std.dev=0.94

kolmogorov-smirnov d=0.0450021 p=n.s

Histogram

NR=1000 uniform (10,20) n=20

bootstrapping for median of n=20



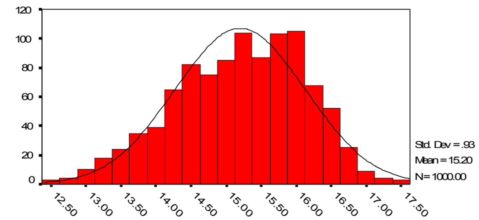
theoritical mean=15 theoritical std. dev=0.81

kolmogorov-smirnov d=0.0417707 p=n.s

Histogram 29

NR=1000 uniform (10,20) n=30

bootstrapping for median of n=30



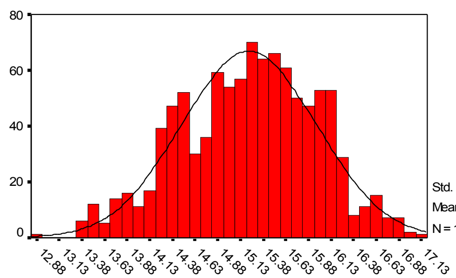
theoritical mean=15 theoritical std. dev=0.66

kolmogorov-smirnov d=0.0474225 p=n.s

Histogram 30

NR=1000 uniform (10,20) n=50

bootstrapping for median of n=50



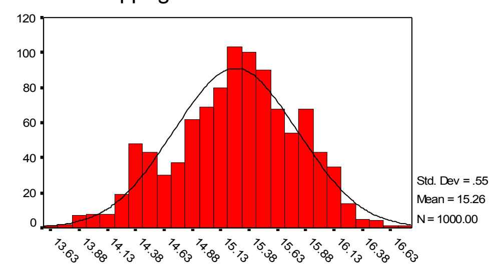
theoritical mean=15 theoritical std. dev=0.52

kolmogorov-smirnov d=0.0330660 p=n.s

Histogram 31

NR=1000 uniform (10,20) n=100

bootstrapping for median of n=100

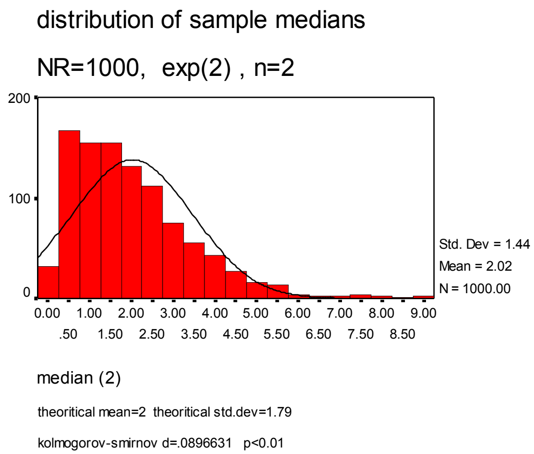


theoritical mean=15 theoritical std. dev=0.36

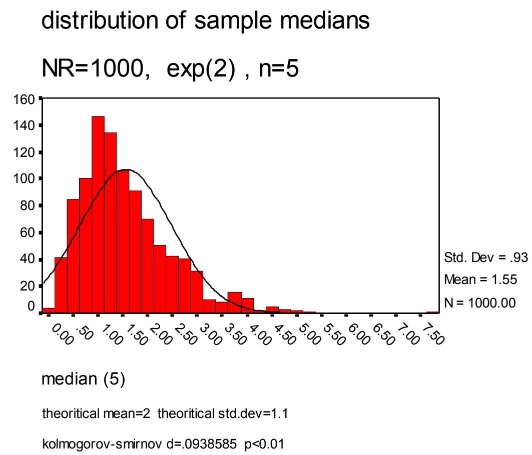
kolmogorov-smirnov d=0.0388305 p=n.s

Histogram 32

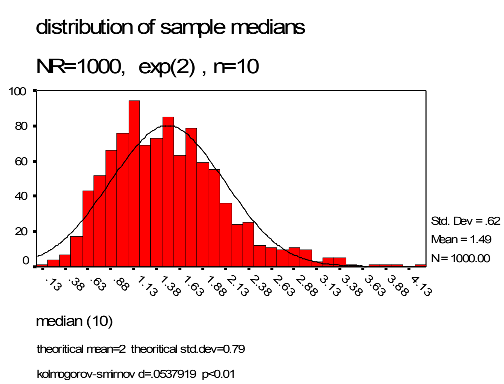
Figure (3.2.5) Probabilistic simulation histograms for the sampling distribution of the sample median. EXP (2).



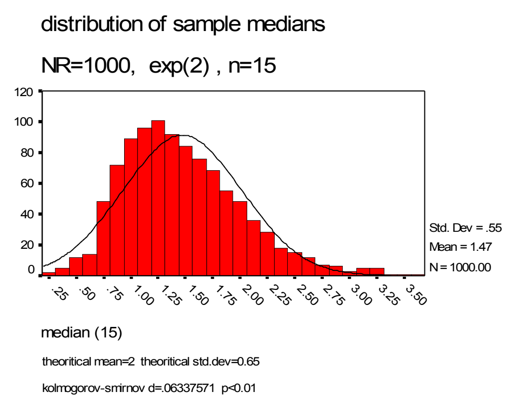
Histogram 33



Histogram 34



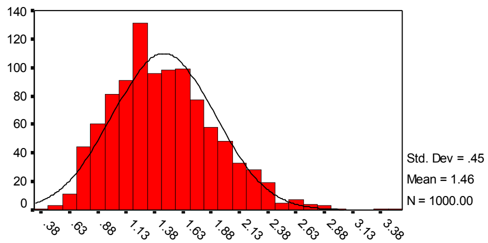
Histogram 35



Histogram 36

distribution of sample medians

NR=1000, exp(2) , n=20



median (20)

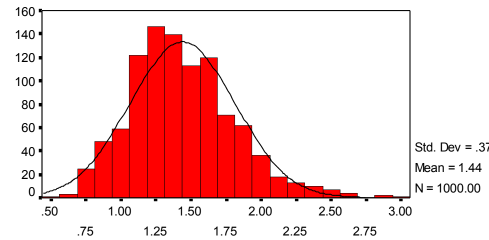
theoritical mean=2 theoritical std.dev=0.56

kolmogorov-smirnov d=.0575298 p<0.01

Histogram 37

distribution of sample medians

NR=1000, exp(2) , n=30



median (30)

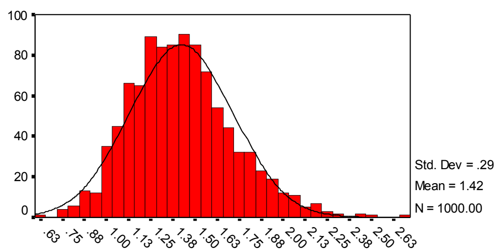
theoritical mean=2 theoritical std.dev=0.46

kolmogorov-smirnov d=.0480907 p<0.01

Histogram 38

distribution of sample medians

NR=1000, exp(2) , n=50



median (50)

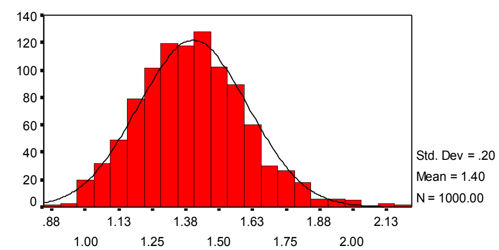
theoritical mean=2 theoritical std.dev=0.35

kolmogorov-smirnov d=.0397730 p<0.01

Histogram 39

distribution of sample medians

NR=1000, exp(2) , n=100



median (100)

theoritical mean=2 theoritical std.dev=0.25

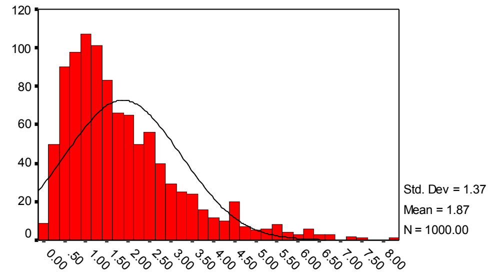
kolmogorov-smirnov d=.02582038 p=n.s

Histogram 40

Figure (3.2.6) Bootstrap simulation histograms for sampling distribution of the sample median, EXP (2).

NR=1000 exp (2) n=2

bootstrapping for median of n=2



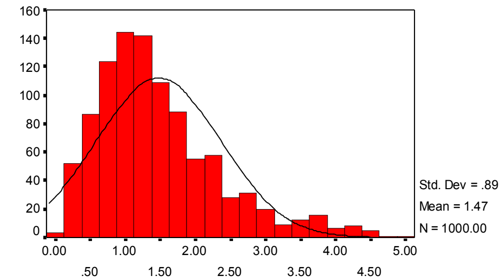
theoritical mean=2 theoritical std.dev=1.77

kolmogorov-smirnov d=0.1114953 p<0.01

Histogram 41

NR=1000 exp (2) n=5

bootstrapping for median of n=5



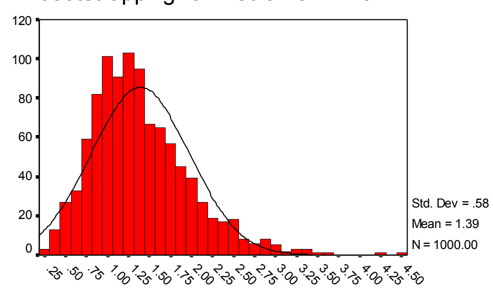
theoritical mean=2 theoritical std.dev=1.12

kolmogorov-smirnov d=0.1085563 p<0.01

Histogram 42

NR=1000 exp (2) n=10

bootstrapping for median of n=10



theoritical mean=2 theoritical std.dev=0.79

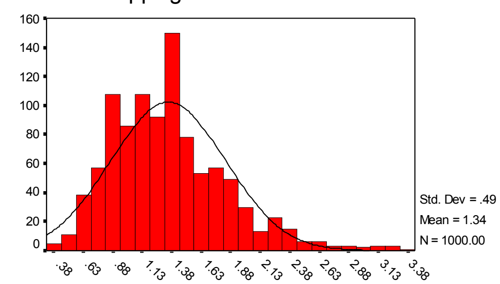
kolmogorov-smirnov d=0.0753861 p<0.01

Histogram 43

Histogram 130

NR=1000 exp (2) n=15

bootstrapping for median of n=15



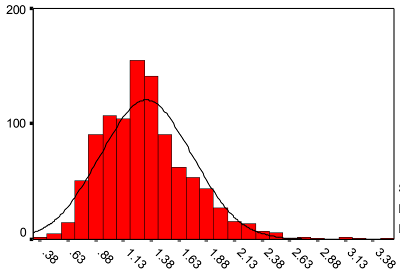
theoritical mean=2 theoritical std.dev=0.65

kolmogorov-smirnov d=0.0836024 p<0.01

Histogram 44

NR=1000 exp (2) n=20

bootstrapping for median of n=20



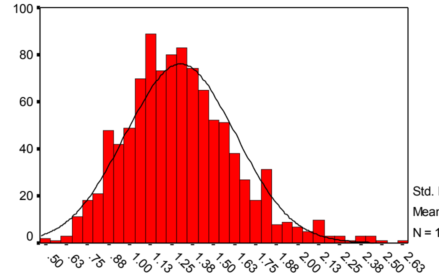
theoritical mean=2 theoritical std.dev=0.56

kolmogorov-smirnov d=0.0716108 p<0.01

Histogram 45

NR=1000 exp (2) n=30

bootstrapping for median of n=30



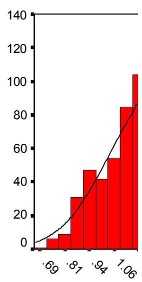
theoritical mean=2 theoritical std.dev=0.46

kolmogorov-smirnov d=0.0464870 p<0.01

Histogram 46

NR=1000

bootstrapp

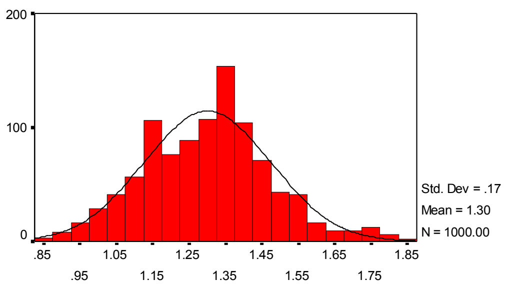


theoritical mean=2

kolmogorov-smirno

NR=1000 exp (2) n=100

bootstrapping for median of n=100



theoritical mean=2 theoritical std.dev=0.25

kolmogorov-smirnov d=0.0304880 p=n.s

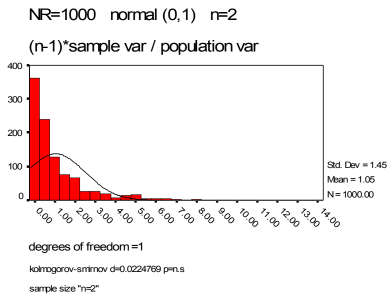
Histogram 47

Histogram 48

APPENDIX 3.

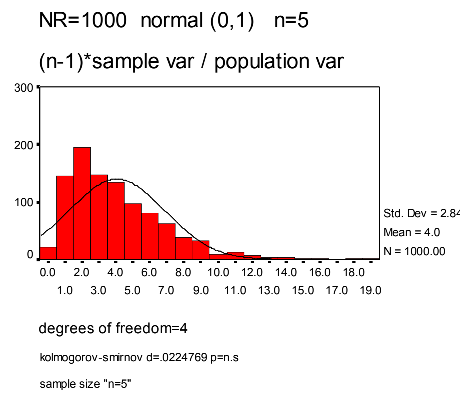
Figure (3.3.1) Probabilistic simulation histograms for sampling distribution of the sample variance, N (0.1)

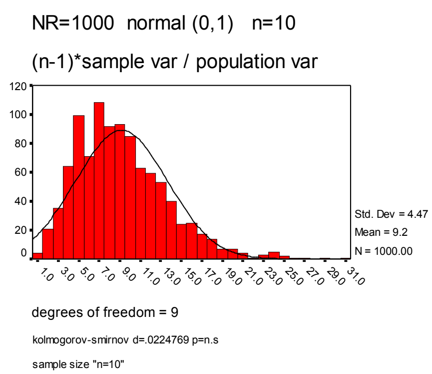
.



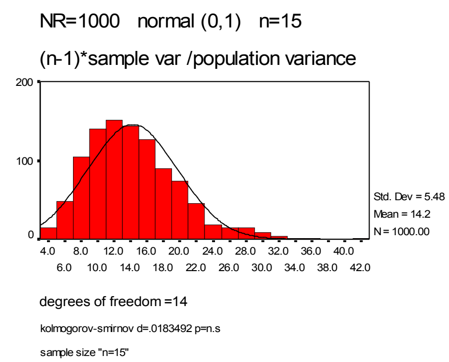
Histogram 1

Histogram 2





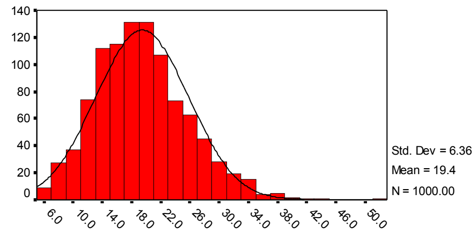
Histogram 3



Histogram 4

NR=1000 normal (0,1) n=20

(n-1)*sample var /population var



degrees of freedom =19

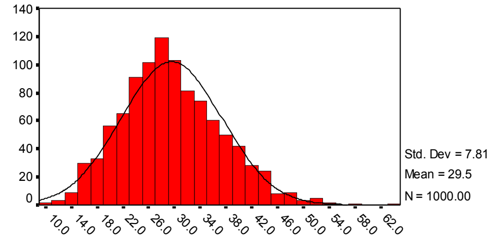
kolmogorov-smirnov d=.0200673 p=n.s

sample size "n=20"

Histogram 5

NR=1000 normal (0,1) n=30

(n-1)*sample var /population var



degrees of freedom =29

kolmogorov-smirnov d=.0148318 p=n.s

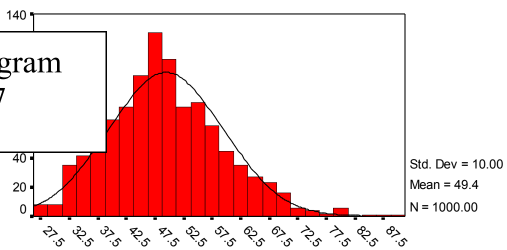
sample size "n=30"

Histogram 6

NR=1000 normal (0,1) n=50

(n-1)*sample var /population var

Histogram 7



degrees of freedom =49

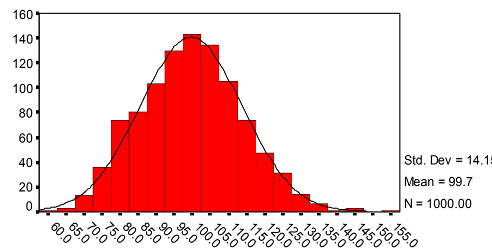
kolmogorov-smirnov d=.0244664 p=n.s

sample size "n=50"

Histogram 8

NR=1000 normal(0,1) n=100

(n-1)*sample var /population var

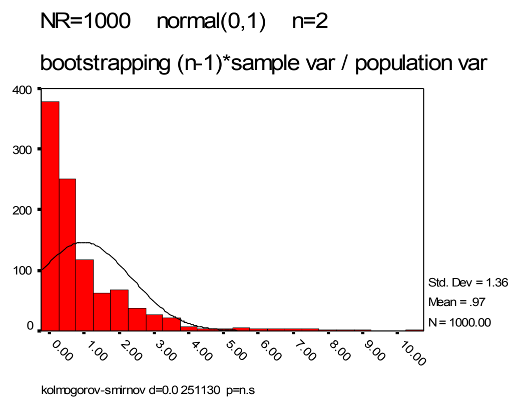


degrees of freedom =99

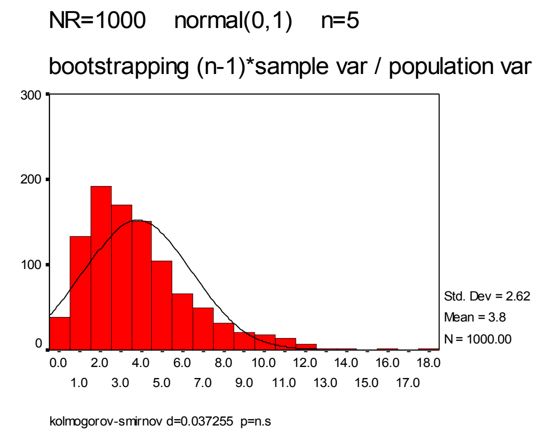
kolmogorov-smirnov d=.0265951 p=n.s

sample size "n=100"

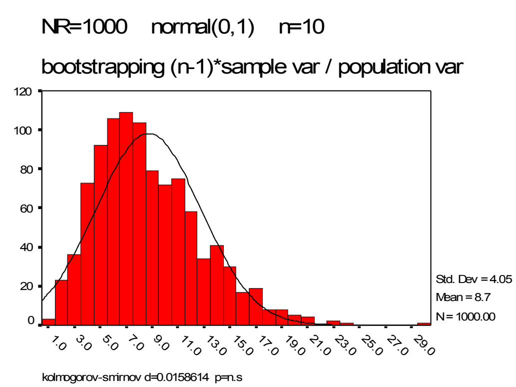
Figure (3.3.20 Bootstrap simulation histograms for the sampling distribution of the sample variance, N (0.1)



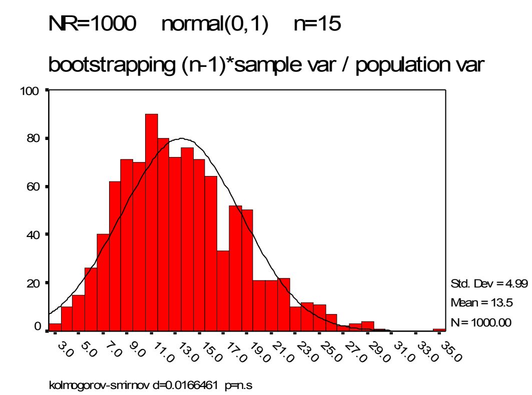
Histogram 9



Histogram 10



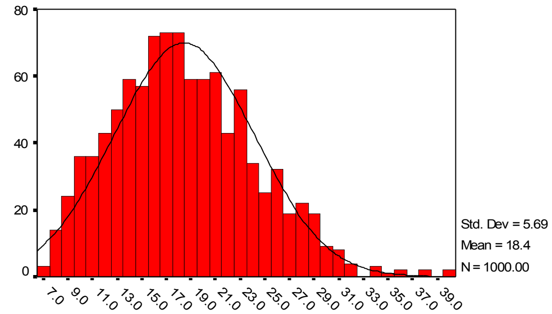
Histogram 11



Histogram 12

NR=1000 normal(0,1) n=20

bootstrapping (n-1)*sample var / population var

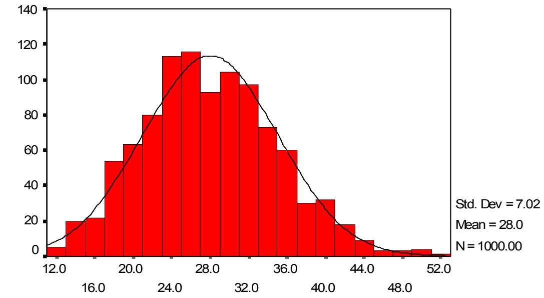


kolmogorov-smirnov d=0.0266919 p=n.s

Histogram 13

NR=1000 normal(0,1) n=30

bootstrapping (n-1)*sample var / population var

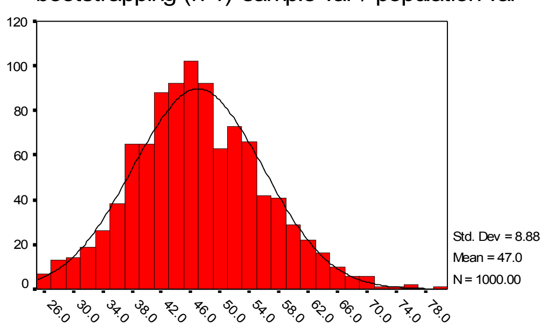


kolmogorov-smirnov d=0.0347390 p<0.01

Histogram 14

NR=1000 normal(0,1) n=50

bootstrapping (n-1)*sample var / population var

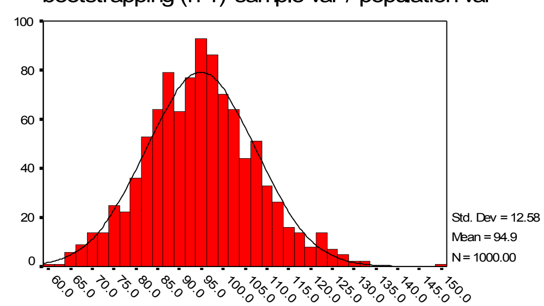


kolmogorov-smirnov d=0.0459341 p<0.05

Histogram 15

NR=1000 normal(0,1) n=100

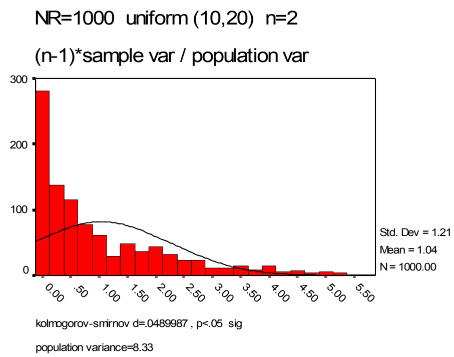
bootstrapping (n-1)*sample var / population var



kolmogorov-smirnov d=0.0377497 p<0.01

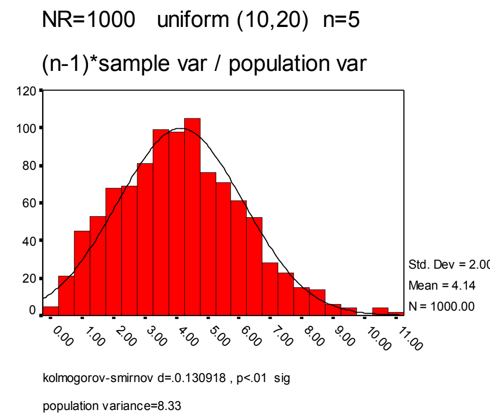
Histogram 16

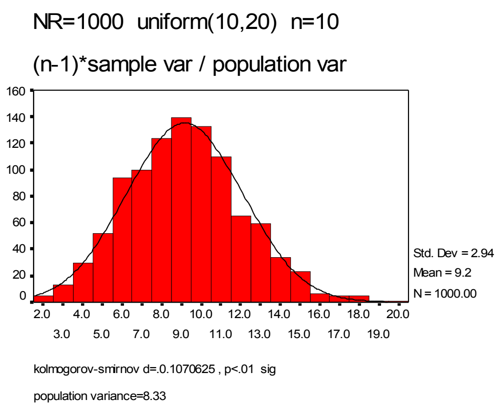
Figure (3.3.3) Probabilistic simulation histograms for the sampling distribution of the sample variance, U (10, 20).



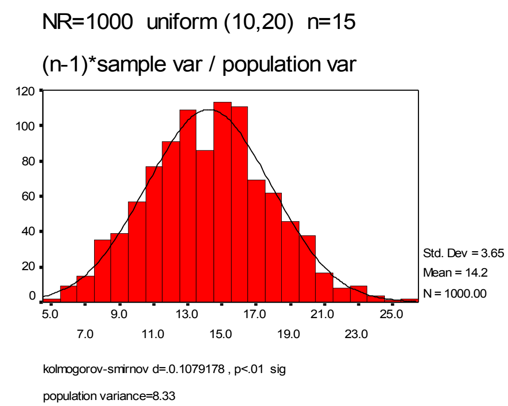
Histogram 17

Histogram 18





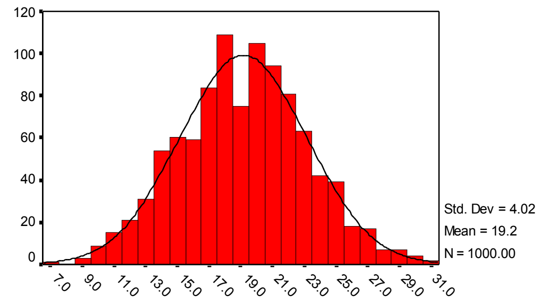
Histogram 19



Histogram 20

NR=1000 uniform (10,20) n=20

(n-1)*sample var / population var



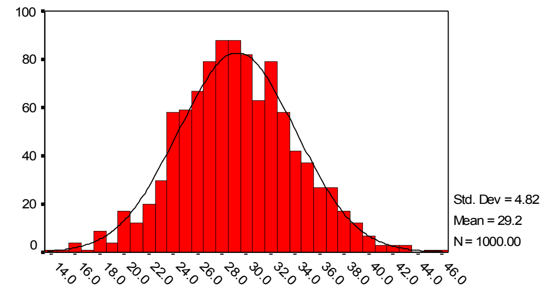
kolmogorov-smirnov d=.0.1193083 , p<.01 sig

population variance=8.33

Histogram 21

NR=1000 uniform(10,20) n=30

(n-1)*sample var / population var



kolmogorov-smirnov d=.1398639 , p<.01 sig

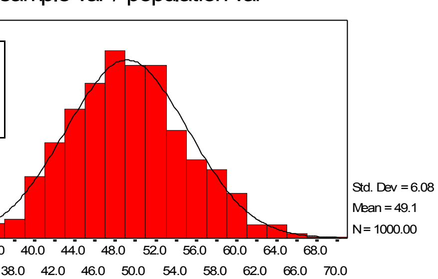
population variance=8.33

Histogram 22

NR=1000 uniform (10,20) n=50

(n-1)*sample var / population var

Histogram 23



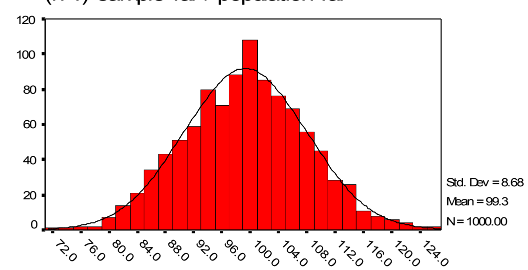
kolmogorov-smirnov d=.1355887 , p<.01 sig

population variance=8.33

Histogram 24

NR=1000 uniform (10,20) n=100

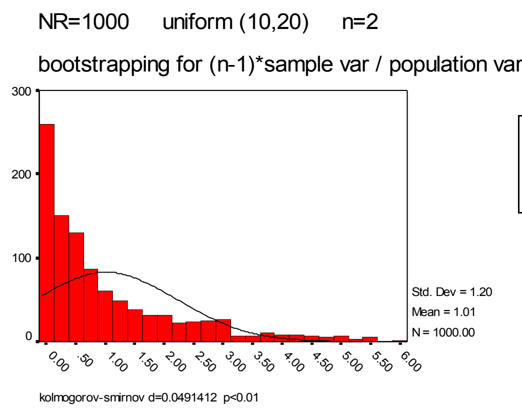
(n-1)*sample var / population var



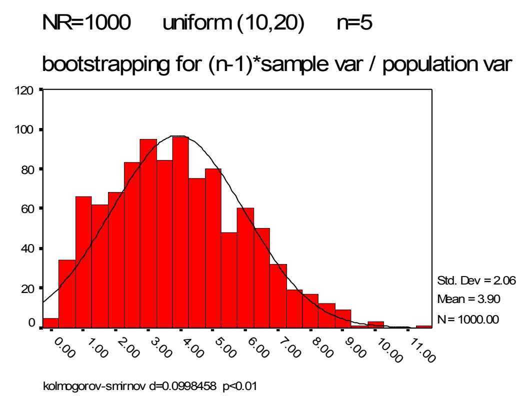
kolmogorov-smirnov d=.1196141 , p<.01 sig

population variance=8.33

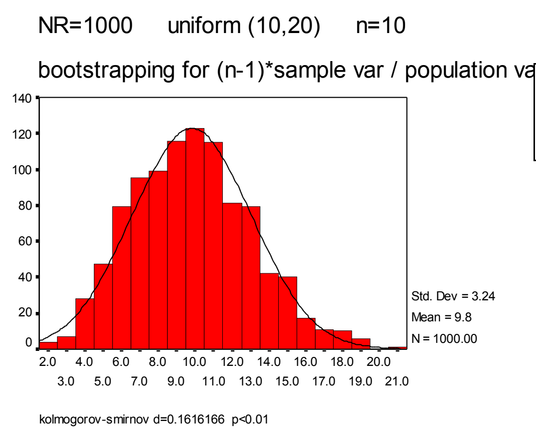
Figure (3.3.4) Bootstrap simulation histograms for the sampling distribution of the sample variance, U (10, 20).



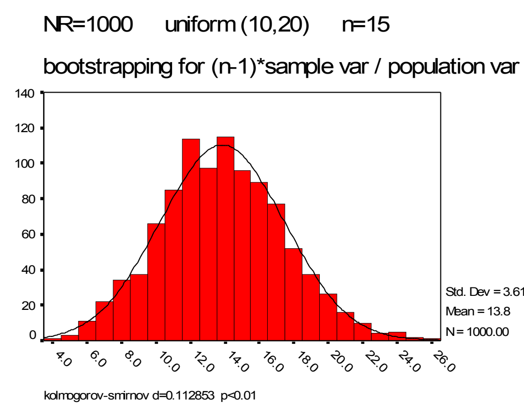
Histogram 25



Histogram 26



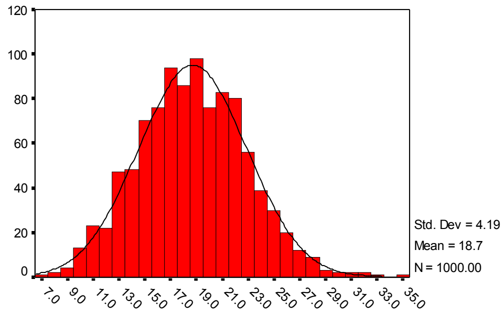
Histogram 27



Histogram 28

NR=1000 uniform (10,20) n=20

bootstrapping for (n-1)*sample var / population var

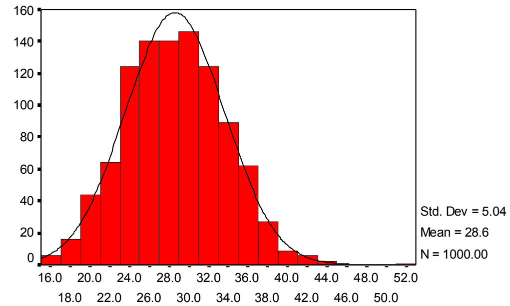


kolmogorov-smirnov d=0.1000362 p<0.01

Histogram 29

NR=1000 uniform (10,20) n=30

bootstrapping for (n-1)*sample var / population var

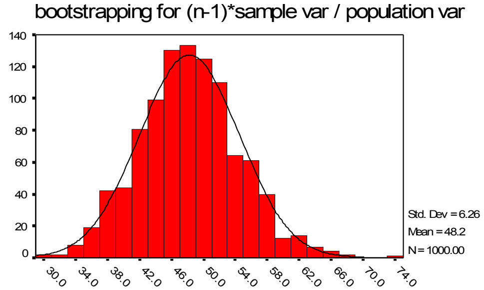


kolmogorov-smirnov d=0.1157174 p<0.01

Histogram 30

NR=1000 uniform (10,20) n=50

bootstrapping for (n-1)*sample var / population var

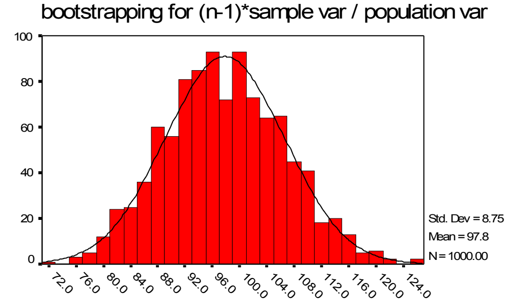


kolmogorov-smirnov d=0.1242285 p<0.01

Histogram 31

NR=1000 uniform (10,20) n=100

bootstrapping for (n-1)*sample var / population var



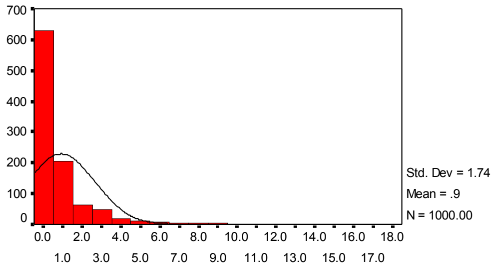
kolmogorov-smirnov d=0.1259373 p<0.01

Histogram 32

Figure (2.3.4) Probabilistic simulation histograms for the sampling distribution of the sample variance, EXP (2).

(n-1)*sample var/population var

NR=1000 exp(2) n=2



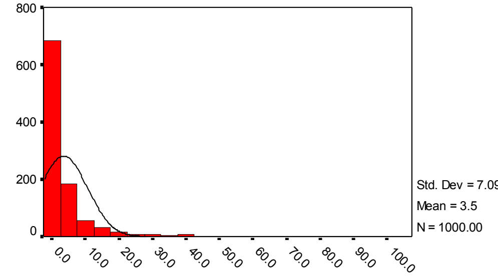
theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.1249231 p<.o1

Histogram 33

(n-1)*sample var / population var

NR=1000 exp(2) n=5



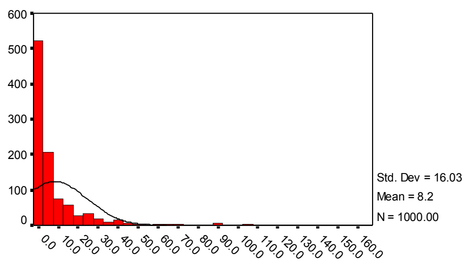
theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.3699096 p<.o1

Histogram 34

(n-1)*sample var / population var

NR=1000 exp(2) n=10



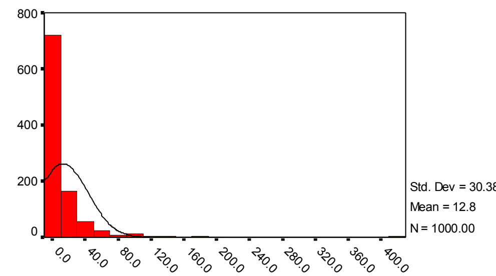
theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.5683179 p<.o1

Histogram 35

(n-1)*sample var / population var

NR=1000 exp(2) n=15



theoritical mean =2 theoritical variance=4

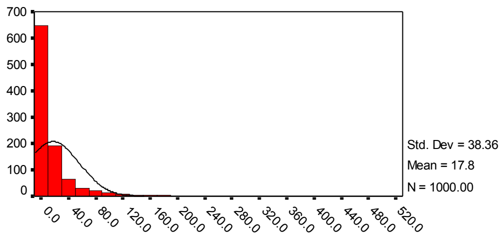
kolmogorov-smirnov d=.5683179 p<.o1

Histogram 36

(n-1)*sample var /population var

NR=1000 exp(2) n=20

sample size "n=20"



Histogram 37

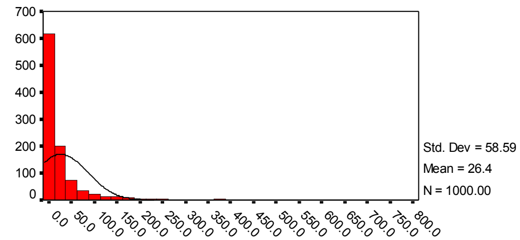
theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.5858901 p<.o1

(n-1)*sample var / pop var

NR=100 exp(2) n=30

sample size "n=30"



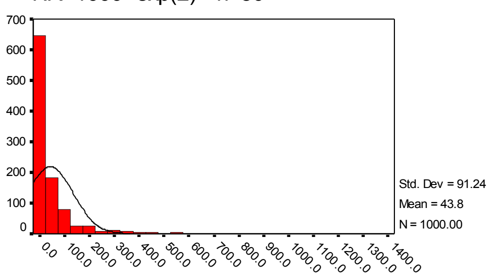
Histogram 38

theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.6263877 p<.o1

(n-1)*sample var / population var

NR=1000 exp(2) n=50



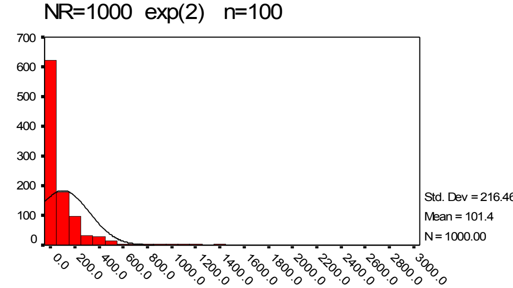
Histogram 39

theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.6425608 p<.o1

(n-1)*sample var / population var

NR=1000 exp(2) n=100



Histogram 40

theoritical mean =2 theoritical variance=4

kolmogorov-smirnov d=.6810809 p<.o1

Figure (3.3.5) Bootstrap simulation for the sampling distribution of the sample variance, EXP (2).