4. Estimation

If the population of data becomes very large, it is difficult to obtain the characteristic values such as the mean or variance. In this case, estimating the characteristic of the population by taking an appropriate number of samples from them is called a statistical estimation.

Making rational decision in uncertain situation using this estimate is an important function of modern statistics.

4.1 Population and Sample

☞ Think	A factory produces thousands of smartphones a day. If the tempered glass on the smartphone screen withstands the impact when it drops from a certain height and does not break, it passes the impact quality test.
Exploration	1) Will this factory conduct an impact test on all smartphones produced? 2) If it is difficult to investigate all smartphones, how should the impact quality test be conducted?

- The quality test for smartphones which may cause a defective or breakage in products cannot be conducted on all products. Therefore, a method of selecting and investigating some of the products produced is usually chosen.

- The entire data subject to a statistical survey are called a **population**, and some of the population selected for the survey are called **samples**. In general, the population is very large.

- In order to find out the characteristic of the population such as the mean, if you survey the entire population, it is called a **total survey**. However, a total survey of all populations requires enormous costs and time. In addition, a total survey is not appropriate if the product can be destroyed during the inspection such as the impact quality test.

- When such a total survey is difficult, a **sample survey** that investigates some samples from the population is used. Estimating the characteristic of the population using samples selected from the population is called an **inferential statistics**.

Example 4.1	Which survey method is more appropriate, a total survey or a sample survey, to investigate the followings? 1) The population born in Korea in 2020. 2) The approval rating of a particular candidate in a parliamentary election.
Solution	1) Since we need to know the exact value of the total number of baby born in 2020 although it might take time and costs, a total survey is more appropriate. 2) It takes a lot of time and effort to investigate the population to know the approval rating of a particular candidate in a parliamentary election. Also, the approval rating of a particular candidate should be known as soon as possible to establish an election strategy based on this result. Therefore, a sample survey is more appropriate.

Exercise 4.1	Which method is more appropriate, a full survey or a sample survey, to investigate the followings? 1) The number of industrial accidents in 2020. 2) The rating of a TV program.

A. Sampling Method

☞ Think	It was decided that the impact quality test of the smartphone is a sample survey.
Exploration	How should we select samples from the large number of smartphones produced per day?

- When we estimate the characteristic of the population using samples, the result may vary depending on how samples are selected, so many studies have been conducted on the sampling method. In general, the most commonly used method for sampling is a **simple random sampling** that ensures that each element in the population has the same probability of being selected. Drawing lots which is often used among friends is an example of a simple random sampling method.

> ☞ Simple random sampling
>
> Each element in the population has the same probability of being selected.

- When samples are selected from a population, there are two types of methods such that one is 'with replacement' and the other is 'without replacement'. After a sample is selected, it is included in the population again in case of **'with replacement'** method. The selected sample is not included in the population in case of **'without replacement'**.

- In general, **random numbers** are used a lot to ensure that each element in the population has the same probability of being selected. Random numbers are numbers that are scattered from 0 to 9 without any particular regularity or bias. [Table 4.1] is a part of the random number table which arranges these random numbers into rows and columns.

{Table 4.1] Random number table arranged in rows and columns

rows	columns									
	1		2		3		4		5	
	12345	67890	12345	67890	12345	67890	12345	67890	12345	67890
01	41061	06462	50931	92621	81915	20895	33925	94062	12002	67659 ⋯
02	60838	83273	53038	18042	04516	17562	91464	31690	98696	96376
03	81155	64589	19627	28271	21929	81536	27757	65190	15141	63172
04	67477	62625	94544	97255	05702	33550	71454	51346	17405	00778
05	71496	89506	21862	48619	60194	99177	20767	27125	05676	69068
06	28042	67711	28387	79997	09614	75357	29592	24664	86804	67350
07	27205	72456	45979	62214	72725	47409	65222	42179	19555	06509
08	83476	96770	92824	51706	46173	79393	16136	28240	40273	44724
09	44883	99390	31386	25418	77886	69148	41159	01688	19665	77392
10	70433	28739	88040	20363	81520	33471	56061	33187	11550	38645

- For example, there are 50 students in a class and their identification numbers are 1, 2, … and 50. In order to select three students of them randomly, first select any one row and one column. If the row 3 and column 11 are selected, read down the double digits starting from here which gives 19, 94, 21, and 28. Among them, numbers outside of 1 to 50 are discarded and three students whose numbers are 19, 21, and 28 are selected.

Exercise 4.2	Seven students from our school created a hobby club and suppose each student is numbered from 1 to 7. Select one chairperson of these seven students using a random number in Table 4.1.

- Simple random sampling using a random number table as shown above is not easy to apply when the sample size is large. Recently random numbers generated by computer using a uniform distribution are commonly used. <Figure 4.1> shows 10 random numbers selected from 1 to 100 numbers without replacement using 『eStatH』.

<Figure 4.1> Ten random numbers generated from Uniform(1,100) by using 『eStatH』

- By using 『eStatH』, you can also generate random numbers of integer-type Uniform $[a, b]$ distribution with replacement, real-type Uniform $[a, b]$ distribution, normal distribution $N(\mu, \sigma^2)$, and two-dimensional normal distribution $N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Example 4.2	A class has 50 students. Using 『eStatH』, select three students randomly without replacement. (In practice, it is not necessary to select samples from a small population of 50, but it is an illustrative example.)
Solution	In order to take three samples out of 50 students, first make a list of students and then assign identification numbers such as from 1 to 50. In order to select three students, select the first option which is the integer uniform distribution without replacement in 'Random Number Generator' of 『eStatH』, Enter (1, 50), enter 3 for 'Number of Data n =' and click [Execute] button, then three random numbers appear as in <Figure 4.2>.

Example 4.2 Solution	
	Random Number Generator　　　　　　　　　　　　　　　Menu ◉ Uniform　　　　　　(1 , 50) Integer　(without replacement) ○ Uniform　　　　　　(1 , 100) Integer　(with replacement) ○ Uniform　　　　　　(0 , 1) Real ○ Normal Distribution　(0 , 1) (Mean μ , Std Dev σ) ○ 2D Normal Distribution (0 , 0 , 1 , 1 , 　0.9) 　　　　　　　　　　(Mean μ₁　Mean μ₂　Std Dev σ₁ Std Dev σ₂　　Correlation Coefficient ρ) Execute　Number of Data n = 3 (≤ 100)　Decimal Digit 0 ≤ 0 ≤ 4　Table Save | i	Uniform(1 , 50) Random Number (without replacement)
1	18
2	32
3	47 | <Figure 4.2> Random number generation of Uniform(1,50) distribution using 『eStatH』

Exercise 4.3	
[QR code]	A class has 30 students. Using 『eStatH』, select five students randomly without replacement.

4.2 Population Mean and Sample Mean

☞ Think	We would like to investigate the height of all first-year high school male students.
Exploration	What is the relationship between the population mean of the height of the entire first-year high school male students and the sample mean of the height of 100 students from this population?

- The purpose of statistical experiments or surveys is to find out the characteristics of the population. If X is a random variable which represents the characteristic of the population, the mean, variance, and standard deviation of the population are called the **population mean, population variance and population standard deviation** denoted by symbols such as μ, σ^2 and σ respectively.

- When n samples are randomly selected from a population, their mean, variance, and standard deviation are called as the **sample mean, sample variance, and sample standard deviation** denoted by symbols such as $\overline{X}, S^2$ and S respectively. If $X_1, X_2, ..., X_n$ represent characteristic values of samples, $\overline{X}, S^2$ and S are defined as follows:

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

$$S^2 = \frac{1}{n-1}\{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + ... + (X_n - \overline{X})^2\}$$

$$S = \sqrt{S^2}$$

In the definition of the sample variance S^2, division by $n-1$ is to estimate the population variance accurately, but its detail explanation can be done in university level of statistics.

- The actual observed values of the random variable $X_1, X_2, ..., X_n$ are denoted as $x_1, x_2, ..., x_n$ and their sample mean, sample variance and sample standard deviation are as follows:

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

$$s^2 = \frac{1}{n-1}\{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2\}$$

$$s = \sqrt{s^2}$$

- For example, when three samples are taken from the first-year high school male population, the possible number of three samples is countless, so this is indicated as random variables X_1, X_2 and X_3. If the actual heights of the three selected students are 160cm, 170cm, and 180cm, these are denoted as $x_1 = 160, x_2 = 170, x_3 = 180$ and their sample mean, sample variance and sample standard deviations are as follows:

$$\overline{x} = \frac{1}{3}(160 + 170 + 180) = \frac{510}{3} = 170$$

$$s^2 = \frac{1}{3-1}\{(160-170)^2 + (170-170)^2 + (180-170)^2\}$$

$$= \frac{100 + 0 + 100}{2} = 100$$

$$s = \sqrt{100} = 10$$

- <Figure 4.3> is a simulation that shows the relationship between the characteristic value of the population and the characteristic value of the sample when about 10% of the sample is randomly selected from 10,000 populations of a standard normal distribution using 『eStatH』 . Looking at

the figure, it can be seen that the minimum (min) and maximum (max) of the population differ from the sample, but there is no significant difference between the population mean and the sample mean.
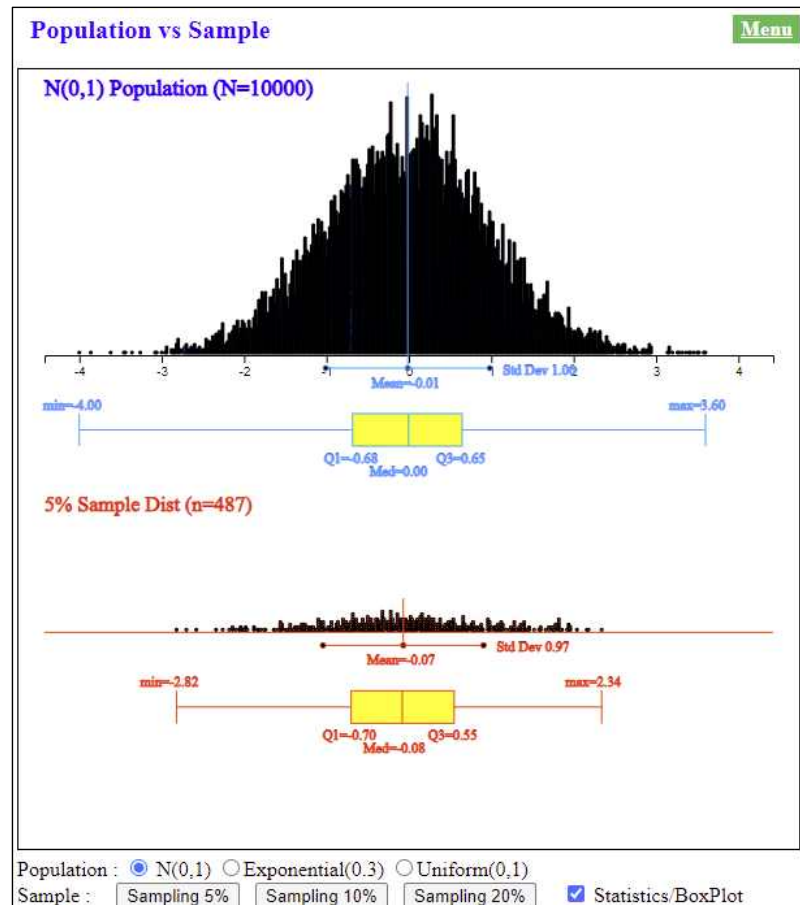
<Figure 4.3> Simulation to show relation between population and sample using 『eStatH』

A. Distribution of all possible sample means

|	
☞ Think	There are five salesman of a company, and the length of time they worked for the company is as follows: 6, 2, 4, 8, 10 (unit year). Suppose that these five employees are a population. (This small population doesn't really need to be sampled, but this is an example to illustrate the distribution of sample means.)
Exploration	If we select all possible samples of size 2 with replacement from the population, what is the relationship between the population mean and all possible sample means of each sample?

• The population mean, variance and standard deviation are as follows:

$$\mu = \frac{1}{5}(6+2+4+8+10) = \frac{30}{5} = 6$$

$$\sigma^2 = \frac{1}{5-1}\{(6-6)^2+(2-6)^2+(4-6)^2+(8-6)^2+(10-6)^2\}$$

$$= \frac{1}{4}(0+16+4+4+16) = \frac{40}{4} = 10$$

$$\sigma = \sqrt{10}$$

The distribution of the population is the unform distribution as follows:
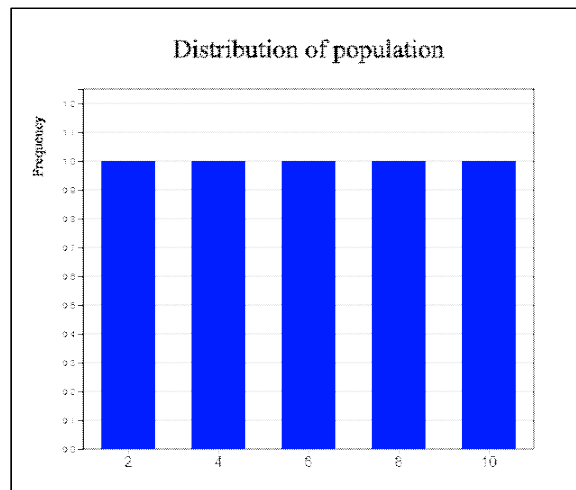
<Figure 4.4> Population distribution

• If we select all possible samples of size 2 with replacement and denote two samples as the random variables X_1 and X_2, the sample mean is a random variable $\overline{X} = \frac{X_1+X_2}{2}$ and a possible value of the sample mean of $\overline{X}$ is $\overline{x}$. The number of all possible sample means of size 2 with replacement is $5^2 = 25$ and each sample mean of the samples, $\overline{x}$, is as follows:

[Table 4.2] All possible samples of size 2 and their sample means from the population

sample	$\overline{x}$	sample	$\overline{x}$	sample	$\overline{x}$	sample	$\overline{x}$	sample	$\overline{x}$
2, 2	2	4, 2	3	6, 2	4	8, 2	5	10, 2	6
2, 4	3	4, 4	4	6, 4	5	8, 4	6	10, 4	7
2, 6	4	4, 6	5	6, 6	6	8, 6	7	10, 6	8
2, 8	5	4, 8	6	6, 8	7	8, 8	8	10, 8	9
2, 10	6	4, 10	7	6, 10	8	8, 10	9	10, 10	10

The frequency table of these 25 sample means and its bar graph are as follows:

[Table 4.3] Frequency table of all possible sample means

sample mean	frequency	relative frequency
2	1	0.04
3	2	0.08
4	3	0.12
5	4	0.16
6	5	0.20
7	4	0.16
8	3	0.12
9	2	0.08
10	1	0.04
	25	1.00
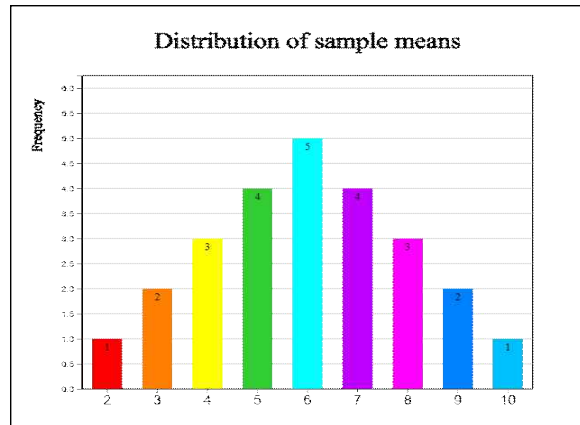
<Figure 4.5> Distribution of all possible sample means

- Looking at the frequency table of all possible sample means in [Table 4.3], some of the sample means exactly match the population mean of 6, and others differ greatly, such as 2 or 10. However, looking at <Figure 4.5>, it can be seen that these sample means are concentrated around the population mean of 6.

- The average of the 25 sample means in {Table 4.2], $E(\overline{X})$, and variance, $V(\overline{X})$, are as follows:

$$E(\overline{X}) = \frac{2 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 4 + 8 \times 3 + 9 \times 2 + 10}{25} = 6$$

$$V(\overline{X}) = \frac{(2-6)^2 + (3-6)^2 \times 2 + (4-6)^2 \times 3 + (5-6)^2 \times 4 + (6-6)^2 \times 5 + (7-6)^2 \times 4 + (8-6)^2 \times 3 + (9-6)^2 \times 2 + (10-6)^2}{25}$$
$$= 4$$

- There are three things that can be observed here.

  - The first is that the mean of all 25 possible sample means is equal to the population mean. This fact is called that the sample mean is the **unbiased estimate** of the population mean.
  - The second is that the variance of all 25 possible sample means is the population variance divided by the size of the sample ($n=2$).
  - Third, when comparing the distribution of the population and the distribution of all possible sample means, the sample means are concentrated around the population mean of 6 and is symmetrical around the population mean.

  - This relationship between the population mean and all possible sample means is observed even if the population is large or has a different distribution form.

☞ **Mean, variance and standard deviation of all possible sample means**

When a population has the mean of μ, the variance of σ^2, if you select samples of size n randomly with replacement, the mean, variance and standard deviation of the sample mean $\overline{X}$ are as follows:

1) $E(\overline{X}) = \mu$

2) $V(\overline{X}) = \dfrac{\sigma^2}{n}$

3) $\sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$

Exercise 4.4	When you select randomly two of the numbers from 1 to 6, find the distribution of the sample mean $\overline{X}$ and find its mean, variance and standard deviation.

Exercise 4.5	When you select randomly samples of 100 people from a population with the mean height of 170cm and population standard deviation of 10cm, find the mean, variance, and standard deviation of the sample mean $\overline{X}$.

- If a population follows the normal distribution of $N(\mu, \sigma^2)$, the distribution of all possible samples of size n which are randomly selected with replacement, the distribution of the sample mean $\overline{X}$ follows the normal distribution of $N(\mu, \dfrac{\sigma^2}{n})$. When the distribution of the population does not follow a normal distribution, if the sample size n is sufficiently large, the distribution of the sample mean $\overline{X}$ follows approximately the normal distribution of $N(\mu, \dfrac{\sigma^2}{n})$.

☞ **Distribution of all possible sample means**

When the population has the mean of μ and variance of σ^2, if you select random samples of size n with replacement, the distribution of sample mean $\overline{X}$ is as follows:

1) If the population follows the normal distribution of $N(\mu, \sigma^2)$, the distribution of the sample mean $\overline{X}$ follows the normal distribution of $N(\mu, \frac{\sigma^2}{n})$.

2) When the population does not follow a normal distribution, if the sample size n is sufficiently large, the distribution of the sample mean $\overline{X}$ follows approximately the normal distribution of $N(\mu, \frac{\sigma^2}{n})$.

- The distribution of the sample mean is very important and it is the fundamental basis of the inferential statistics. <Figure 4.6> is a simulation in 『eStatH』 that shows how the distribution of sample mean varies depending on the size of the sample when the population follows the normal distribution form. Looking at the figure, as the sample size increases, the distribution of the sample mean has the same mean as the population mean, the shape becomes a normal distribution like, and the variance gradually becomes smaller and pointed. This means that if the sample size is very large, the distribution of all possible sample means will be concentrated around the population mean. In other words, the sample mean we obtain is one of all possible sample means, but it is close to the population mean, which is the basis for estimating the population mean well by using any sample mean.
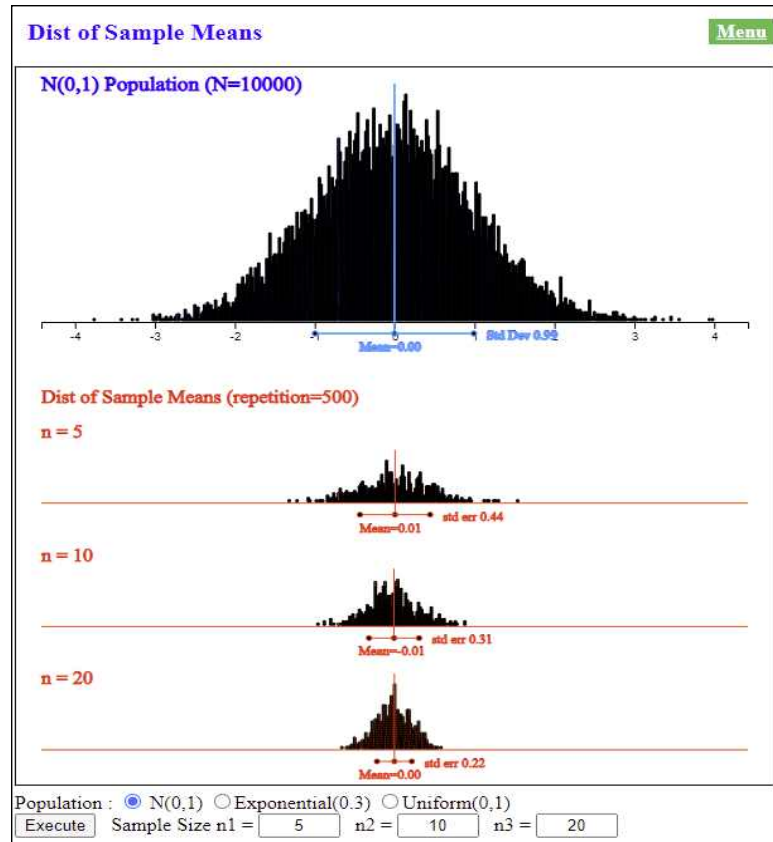
<Figure 4.6> Simulation of possible sample means using 『eStatH』

Example 4.3	It is said that the height of the first-year high school students follows a normal distribution with the mean of 170cm and standard deviation of 10cm. When you select samples of 100 students randomly, find the following probability. 1) $P(168 \leq \overline{X} \leq 172)$ 2) $P(\overline{X} \geq 173)$
Solution	$\overline{X}$ is the random variable with the mean of 170, variance of $\frac{10^2}{100} = 1$, standard deviation of 1 which follows the normal distribution of $N(170, \frac{10^2}{100})$. 1) $P(168 \leq \overline{X} \leq 172) = P(\frac{168-170}{\frac{10}{\sqrt{100}}} \leq \frac{\overline{X}-170}{\frac{10}{\sqrt{100}}} \leq \frac{172-170}{\frac{10}{\sqrt{100}}})$ $\qquad\qquad\qquad\qquad = P(-2 \leq Z \leq 2) = 0.955$ 2) $P(\overline{X} \geq 173) = P(\frac{\overline{X}-170}{\frac{10}{\sqrt{100}}} \geq \frac{173-170}{\frac{10}{\sqrt{100}}}) = P(Z \geq 3) = 0.001$. It shows that the most of the sample mean located in the interval of $170 \pm 2 \times \frac{10}{\sqrt{100}}$, that is [168, 172].

Exercise 4.6	It is said that the length of bolts produced in one factory follows a normal distribution with the mean of 100mm and standard deviation of 2mm. When we select randomly 25 samples with replacement, find the following probability.
	1) $P(99.5 \leq \overline{X} \leq '100.5)$
	2) $P(\overline{X} \geq '100.3)$

4.3 Estimation of Population Mean

☞ Think	It is intended to estimate the average height of all first-year high school mail students. When 10 high school first-year male students were randomly selected and their heights were investigated, the sample mean was 170cm.
Exploration	Can I use the sample mean as an estimate of the mean height of the entire high school first-year male students?

- When a sample survey is conducted, only one set of samples is taken from the population to estimate the population mean. In general, we think the mean of the samples as an estimate of the population mean, but can the sample mean obtained from one set of samples among countless possible set of samples predict the population mean well?

- The sampling distribution of all possible sample means studied in the previous section is the answer to this question. That is, no matter what a population distribution is, if the sample size is large enough, all possible sample means become dense with a normal distribution shape around the population mean. Therefore, the mean of a set of samples we obtained is usually close to the population mean, and even in the worst case, the difference from the population mean (referred to as an **error**) is not large, so the sample mean can be said to be a good estimate of the population mean. The larger the sample size, the more dense the distribution of the sample means is, near the population mean, so this error decreases.

- The fact that one value of the observed sample mean is an estimate of the population mean is called the **point estimation** of the population mean (meaning that it is estimated as one number). The sample mean used to estimate the population mean has several good characteristics. One characteristic examined in the previous section is that the mean of all possible sample means becomes to the population mean, and this

characteristic is called the **unbiased estimate**.

- Unlike the point estimation, estimating the population mean as an interval is called the **interval estimation**. If the population is a normal distribution with mean μ, variance σ^2, the distribution of the sample mean $\overline{X}$ is the normal distribution with mean μ and variance σ^2/n, so the probability that one sample mean will be included in the interval $\mu \pm 1.96\sigma/\sqrt{n}$ is 95% as follows:

$$P(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

This equation can also be written as follows:

$$P(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

- The meaning of this equation is that 95% of all possible intervals obtained by applying the following interval formula (assuming that σ is known) to all possible sample means include the population mean μ. The following interval formula is called the **95% confidence interval** of the population mean.

$$\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} , \overline{X} + 1.96\frac{\sigma}{\sqrt{n}} \right]$$

- In general, since $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, the standardized random variable of $\overline{X}$ that is $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$, follows the standard normal distribution $N(0,1)$. Therefore we can write the following probability statement.

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

The left hand side of the aboe equation can also be written as follows:

$$P(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$(1 - \alpha)$ is called the **confidence level** that refers to the probability of intervals in which the population mean is included among all intervals calculated by this interval formula. Usually, 0.01 or 0.05 is used for α. z_{α} means a value with a probability α at the right end of the standard normal distribution. In other words, if Z is a random variable that follows the

standard normal distribution, the probability Z is greater than z_α is α, and if expressed by an equation, $P(Z \geq z_\alpha) = \alpha$ is established. For example, $z_{0.025} = 1.96$, $z_{0.005} = 2.575$, etc. Therefore, the 99% confidence interval formula of the population mean is as follows:

$$\left[\overline{X} - 2.575 \frac{\sigma}{\sqrt{n}} \ , \ \overline{X} + 2.575 \frac{\sigma}{\sqrt{n}} \right]$$

- <Figure 4.7> shows 95% confidence intervals for the population mean by randomly selecting 100 sets of samples with the sample size $n=20$ from 10,000 population data which follow the standard normal distribution. In this case, only 94 out of 100 confidence intervals contain the population mean of 0. Whenever these experiments are repeated, results can vary slightly. This simulation shows that, when estimating the confidence interval of the population mean several times, the confidence interval by using the above formula includes the population mean approximately to the confidence level.
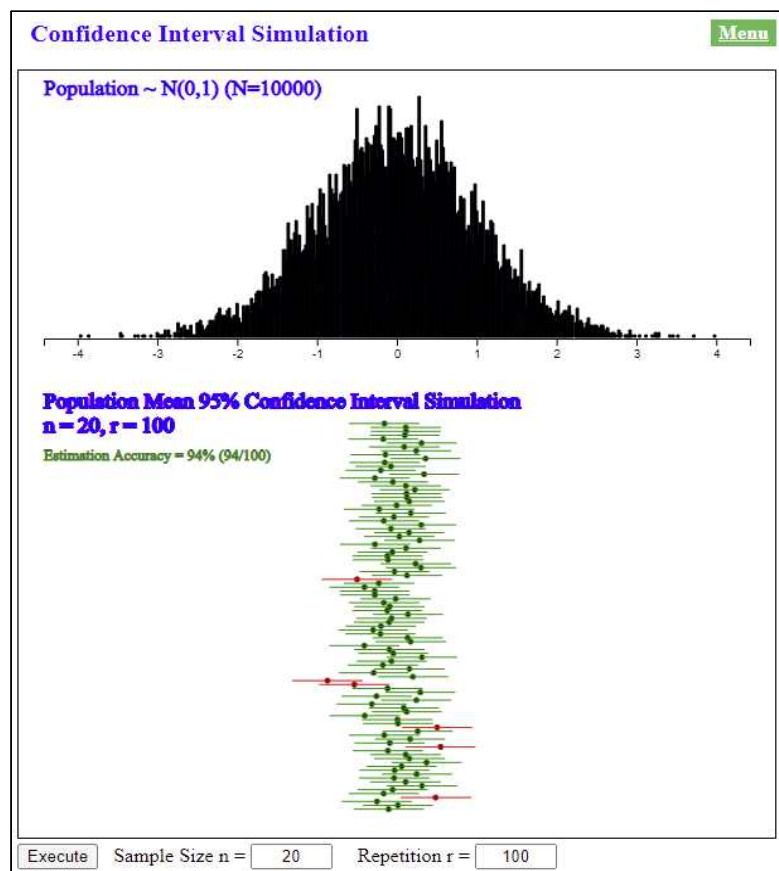
<Figure 4.7> Simulation of 95% confidence interval using 『eStatH』

- When a population follows the normal distribution with the mean μ and variance σ^2 which is known, the confidence interval for the population mean μ with the confidence level of $1-\alpha$ is as follows:

☞ Interval estimation of μ with 100(1−α)% confidence level

When the population follows the normal distribution with the mean μ and variance σ^2 which is known, 100(1−α)% confidence interval for μ is as follows:

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} , \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right]$$

95% confidence interval for μ: $\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} , \overline{X} + 1.96\frac{\sigma}{\sqrt{n}} \right]$

99% confidence interval for μ: $\left[\overline{X} - 2.575\frac{\sigma}{\sqrt{n}} , \overline{X} + 2.575\frac{\sigma}{\sqrt{n}} \right]$

| Example 4.4	The height of all first-year high school male students in Korea follows the normal distribution and the population standard deviation is known to be 10cm. When 100 high school first-year male students were randomly selected and their heights were investigated, the sample mean was 170cm. About the mean height of male students in the first year of high school, answer the followings. 1) Find the 95% confidence interval. 2) Find the 99% confidence interval. 3) When the sample size is 400, if the sample mean is 170cm, find the 95% confidence interval. Compare this confidence interval with the confidence interval in 1).
Solution	1) If the confidence level is 95%, α = 0.05. Therefore, $z_{\alpha/2} = z_{0.05/2} = z_{0.025} = 1.96$ The 95% confidence interval is as follows: $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} , \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right]$ ⇔ [170 − 1.96(10/10), 170 + 1.96(10/10)] ⇔ [168.04, 171.96] 2) If the confidence level is 99%, α = 0.01. Therefore, $z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.575$ The 99% confidence interval is as follows: [170 − 2.575(10/10), 170 + 2.575(10/10)] ⇔ [267.425, 272.575] If the confidence level increases, the interval width widens.

| Example 4.4
Solution 계속 | 3) If the sample size is 400, the 95% confidence interval is as follows:
    [170 − 1.96(10/20) , 170 + 1.96(10/20)]
    ⇔ [169.02, 170.98)
 Therefore, as the sample size increases, the width of the confidence interval narrows. In other words, it is a more accurate estimation. |

Using 『eStatH』, it is easy to estimate the population mean. Select 'Population Mean Confidence Interval' from 『eStatH』 menu, enter 'Sample size', 'Sample Mean', [Population Standard Deviation] and select the confidence level as shown in <Figure 4.8>. If you click [Execute] button, the confidence interval is calculated and the interval is drawn on the graph window. Under the interval graph, there is a simulation window that can adjusts the sample size and confidence level.
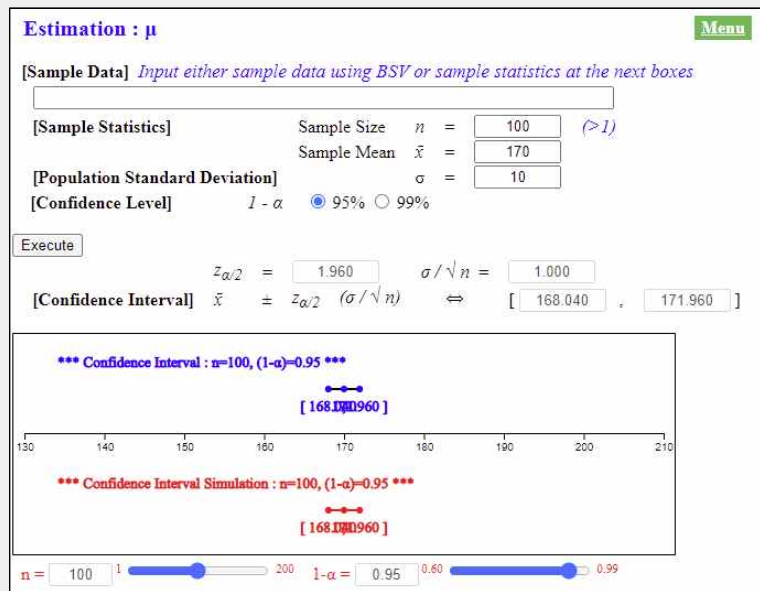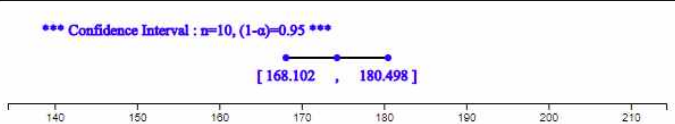
<Figure 4.8> Interval estimation of population mean using sample statistics in 『eStatH』

| Exercise 4.7 | It is said that the length of bolts produced in a factory follows a normal distribution with the population standard deviation of 2mm. If the sample mean is 101mm when 25 samples are randomly selected, answer the followings.
1) Find a 95% confidence interval of the population mean of the bolt length.
2) Find a 99% confidence interval of the population mean of the bolt length.
3) Compare the widths of the two confidence intervals. |

Example 4.5	The height of all first-year high school male students follows the normal distribution with the population standard deviation of 10cm. Ten high school first-year male students were randomly selected and their heights were investigated as follows: 171, 172, 185, 169, 175, 177, 174, 179, 168, 173 Using 『eStatH』, answer the followings about the mean height of male students in the first grade of high school. 1) Find the 95% confidence interval. 2) Find the 99% confidence interval. 3) Compare the above two confidence intervals.
Solution	If there are data of samples, it is easy to estimate the population mean by using 『eStatH』. Select 'Population Mean Confidence Interval' from 『eStatH』 menu, enter the data as shown in Figure 4.9, and click [Execute] button to calculate the 95% confidence interval. <Figure 4.9> Interval estimation of population mean using sample data in 『eStatH』 If you select the 99% confidence level and click [Execute], the 99% confidence interval is calculated. Under the interval graph, there is a simulation window that can adjusts the sample size and confidence level.

Exercise 4.8	It is said that the length of bolts produced in a factory follows the normal distribution with the population standard deviation of 2mm. The data obtained by randomly extracting 20 samples are as follows:
	99 101 102 99 102 101 100 97 99 98 103 97 106 100 99 106 101 98 101 97
	Using 「eStatH」, answer the followings. 1) Find the 95% confidence interval of the mean length of the bolts. 2) Find the 99% confidence interval of the mean length of the bolts. 3) Investigate the change in confidence interval according to the confidence level and sample size.

Exercise

4.1 What is the set of whole data for a characteristic value? (Answer ③)
 ① whole data ② sample
 ③ population ④ subset

4.2 To find out one characteristic of a population, what is a subset of the population called? (Answer ②)
 ① whole data ② sample
 ③ population ④ subset

4.3 Which survey method is more appropriate to investigate the number of foreign workers in one country in 2020?(Answer ②)
 ① sample survey ② total survey
 ③ small survey ④ visiting survey

4.4 Which survey method is more appropriate to investigate the approval rating of a particular candidate in a presidential election? (Answer ①)
 ① sample survey ② total survey
 ③ small survey ④ visiting survey

4.5 A population follows the normal distribution with mean of 10 and variance of 2. If you select 10 samples randomly from the population, what is the average of the sample mean $\overline{X}$? (Answer ④)
 ① 1   ② 2
 ③ 5   ④ 10

4.6 A population follows the normal distribution with mean of 10 and variance of 4. If you select 10 samples randomly from the population, what is the variance of the sample mean $\overline{X}$? (Answer ②)
 ① 4   ② 0.4
 ③ 5   ④ 10

4.7 It is said that the length of bolt products produced in one factory follows a normal distribution with the mean of 10mm and the standard deviation of 1mm. When 25 samples are randomly selected from this population, find $P(9.8 \leq \overline{X} \leq 10.2)$. (Answer ④)
 ① 1.000   ② 0.955
 ③ 0.999   ④ 0.683

4.8 It is said that the length of bolt products produced in one factory follows a normal distribution with the standard deviation of 1mm. When 25

samples are randomly selected from this population, the sample mean is 9mm. What is the 95% confidence interval for the mean length of bolt products? (Answer ②)

① $9 \pm 1.96 \times 1$

② $9 \pm 1.96 \times \dfrac{1}{\sqrt{25}}$

③ $9 \pm 2.575 \times \dfrac{1}{25}$

④ $9 \pm 1.96 \times \dfrac{1}{25}$

Answer
4.1 ③, 4.2 ②, 4.3 ②, 4.4 ①, 4.5 ④, 4.6 ②, 4.7 ④, 4.8 ②