Chapter 6

Sampling Distribution and Estimation

Professor Jung Jin Lee

Table of Contents

Chapter 6 Sampling Distribution and Estimation

6.1 Simple Random Sampling

6.2 Sampling Distribution of Sample Means and Estimation of Population Mean

6.3 Sampling Distribution of Sample Variance and Estimation of Population Variance

6.4 Sampling Distribution of Sample Proportion and Estimation of Population Proportion

6.5 Determination of Sample Size

6.1 Simple Random Sampling

- Since a population is generally very large, statistical surveys of the entire population takes a lot of money and time.
- Estimation the characteristics of the entire population using samples from some of the population is called inferential statistics.
- However, there may be a difference between the characteristics of the population and the findings of the samples.
- To reduce these differences, several methods of sampling have been studied, many of which are simple random sampling which collects a sample with the same probability of all elements of the population being selected.

6.1 Simple Random Sampling

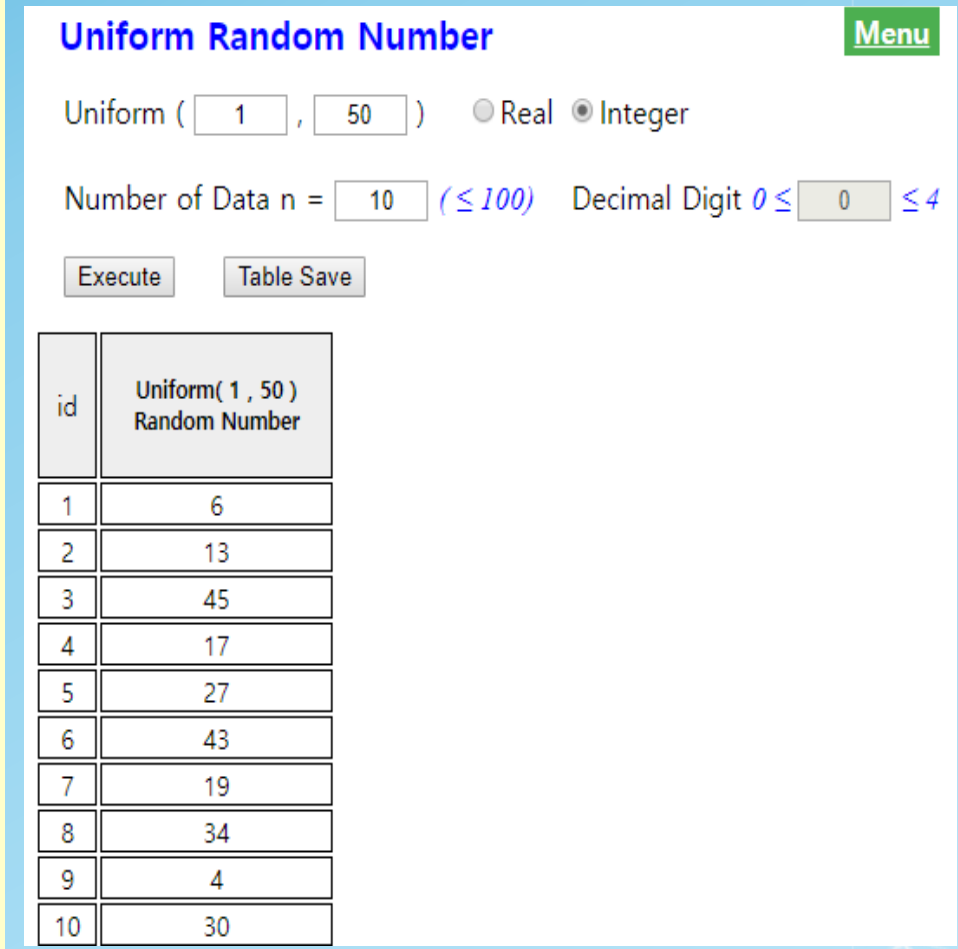
- In case of simple random sampling, it is possible to include elements extracted once again in the population (with replacement), or that does not include the extracted elements back into the population (without replacement). However, in practice, almost all sampling is made without replacement.
- Some tools may be needed to ensure that each element of the population is selected equally. We usually use a random number table. A random number table is a table of numbers from 0 to 9 without special regularity or partiality.
- Recently, a uniform distribution on [0, 1] is used to produce a random number by using the computer.

6.1 Simple Random Sampling

[Ex 6.1.1] A class has 50 students. Select three of these students as a sample without replacement by using 『eStatU』.

〈Answer〉

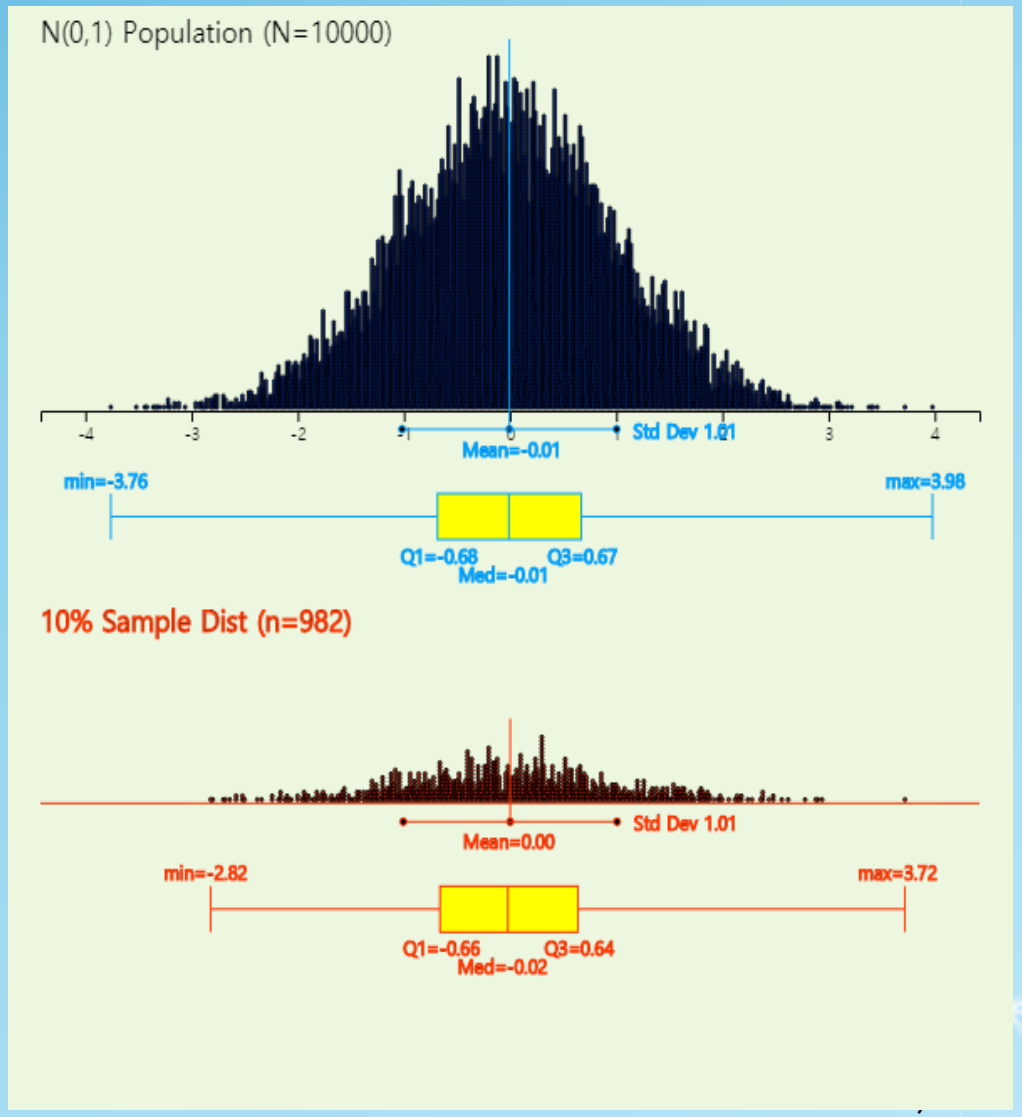
- A student's list must first be made and serial numbers must be assigned from one to fifty.
- To extract students, select 'Uniform Random Number' in 『eStatU』. Enter (1 and 50) at the box of 'Uniform', check 'Integer', enter 10 at the box of 'Number of Data' and click [Execute].
- Since there is no overlap of 6, 13 and 45, and you will choose these student numbers as a sample.
- The generated random number may have the same numbers. If it is a sampling without replacement, discard the following same number.



id	Uniform(1 , 50) Random Number
1	6
2	13
3	45
4	17
5	27
6	43
7	19
8	34
9	4
10	30

6.2 Sampling Distribution of Sample Means and Estimation of Population Mean

- The purpose of statistical experiments is to find out information about unknown populations.
- Information in a population usually refers to the characteristic values of a population, such as means, variance, etc., and the characteristic values of those populations is called parameters.
- The parameters are usually estimated by taking samples such as sample mean and sample variance.



6.2.1 Sampling Distribution of Sample Means

[Ex 6.2.1] Consider a population consisting of five salespeople from a company. Consider a random variable of the number of years of service at this company and the population data is as follows.

6, 2, 4, 8, 10

1) Obtain the mean and variance of this population.
2) Obtain all possible samples of size two by simple random sampling withou replacement in this population and calculate each sample mean. In addition, calculate the mean and variance of all these possible sample means and compare them with the mean and variance of the population.
3) Prepare a frequency distribution of sample means and draw a bar chart. Compare this with the distribution of the population.

6.2.1 Sampling Distribution of Sample Means

⟨Answers of Ex6.2.1⟩

1) The mean and variance of the population is $\mu = 6$, $\sigma^2 = 8$
2) The number of all samples with replacement is $5 \times 5 = 25$. Table 6.2.1 shows all possible samples and their sample means($\bar{x}$).

Table 6.2.1 All possible samples of size 2 from the population and their sample means

sample	$\bar{x}$	sample	$\bar{x}$	sample	$\bar{x}$	sample	$\bar{x}$	sample	$\bar{x}$
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

6.2.1 Sampling Distribution of Sample Means

⟨Answers of Ex6.2.1⟩

- Some of these sample means are exactly the same as the population mean μ = 6, but some others are significantly different such as 2 or 10. The mean of all possible 25 sample means (indicated by $\mu_{\overline{X}}$) in Table 6.2.1 is also 6 and the variance (indicated by $\sigma^2_{\overline{X}}$) is 4 as follows.

$$\mu_{\overline{X}} = \frac{2+3\times2+4\times3+5\times4+6\times5+7\times4+8\times3+9\times2+10}{25} = 6$$

$$\sigma^2_{\overline{X}} = \frac{(2-6)^2+(3-6)^2\times2+(4-6)^2\times3+(5-6)^2\times4+(6-6)^2\times5+(7-6)^2\times4+(8-6)^2\times3+(9-6)^2\times2+(10-6)^2}{25}$$

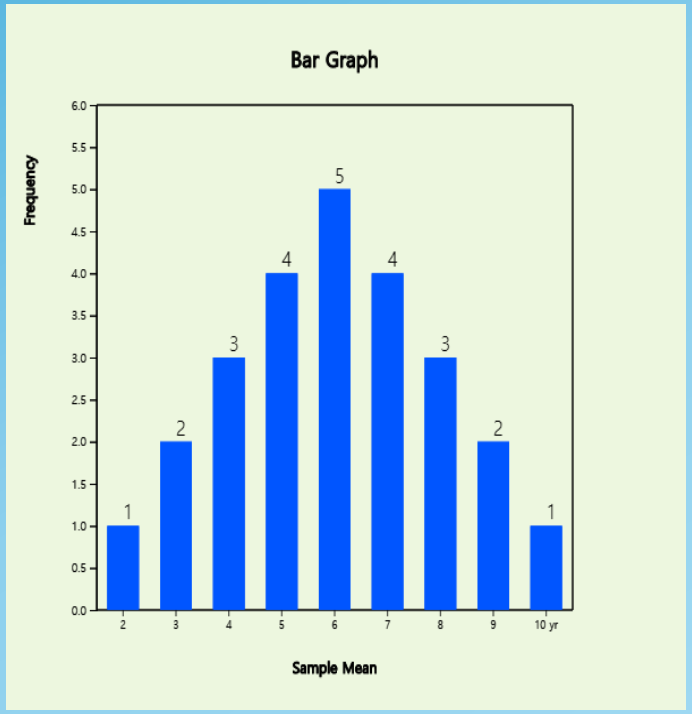
$$= 4$$

- What can be observed here is that the mean of all 25 possible sample means is the same as the population mean. This fact is said that the sample mean $\overline{x}$ is the **unbiased estimator** of the population mean μ. In addition, the variance of the sample means $\sigma^2_{\overline{X}}$ is the population variance σ^2 divided by the size of the sample (n=2).

6.2.1 Sampling Distribution of Sample Means

〈Answers of Ex6.2.1〉
3)

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



- Mean of all 25 possible sample means is the same as the population mean. This fact is said that the sample mea is an unbiased estimator of the population mean
- Variance of the sample means is the population variance divided by the sample size

6.2.1 Sampling Distribution of Sample Means

- If the population is normally distributed $N(\mu, \sigma^2)$, the distribution of all possible sample means is exactly normal distribution $N(\mu, \sigma^2/n)$.
- If the population is an infinite population with mean μ and variance σ^2, then the distribution of all possible sample means is approximately normal distribution $N(\mu, \sigma^2/n)$ when the sample size is large enough. This is referred to as the **Central Limit Theorem**, which is specifically summarized as follows:

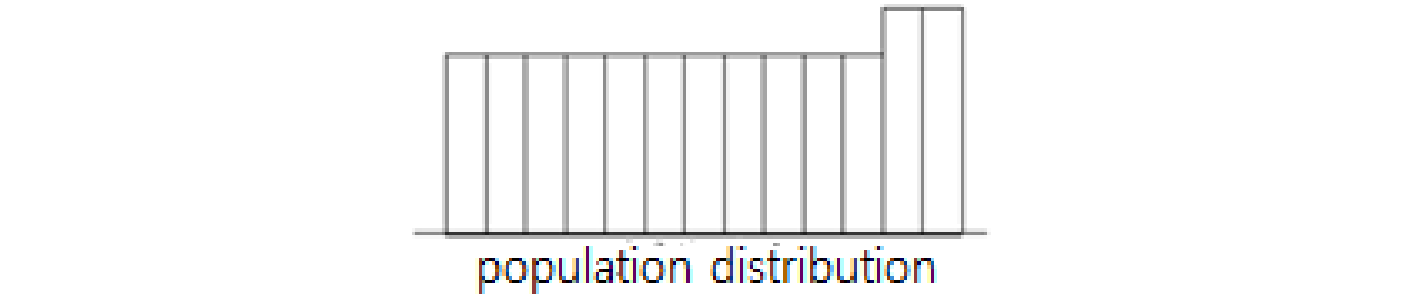
(1) The average of all possible sample means $\mu_{\overline{X}}$ is equal to the population mean μ. (i.e., $\mu_{\overline{X}} = \mu$)

(2) The variance of all possible sample means $\sigma^2_{\overline{X}}$ is the population variance divided by n. (i.e., $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$)

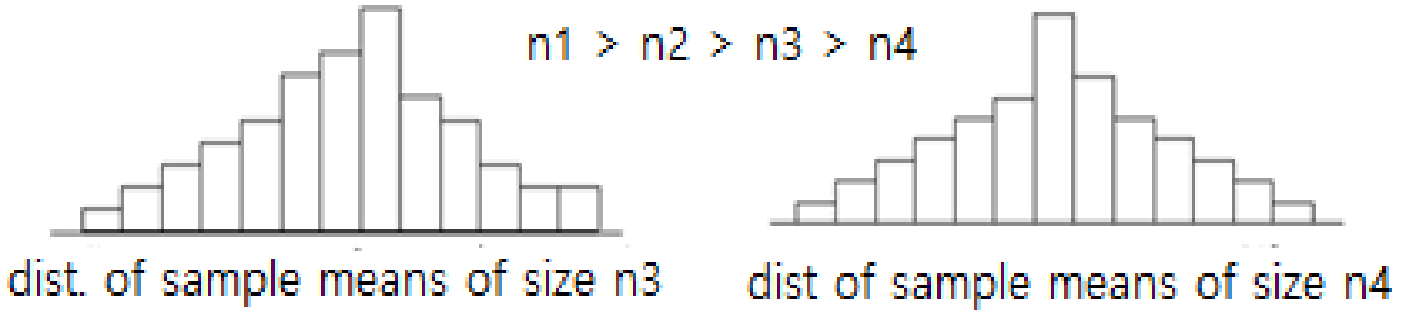
(3) The distribution of all possible sample means is approximately normal.

6.2.1 Sampling Distribution of Sample Means

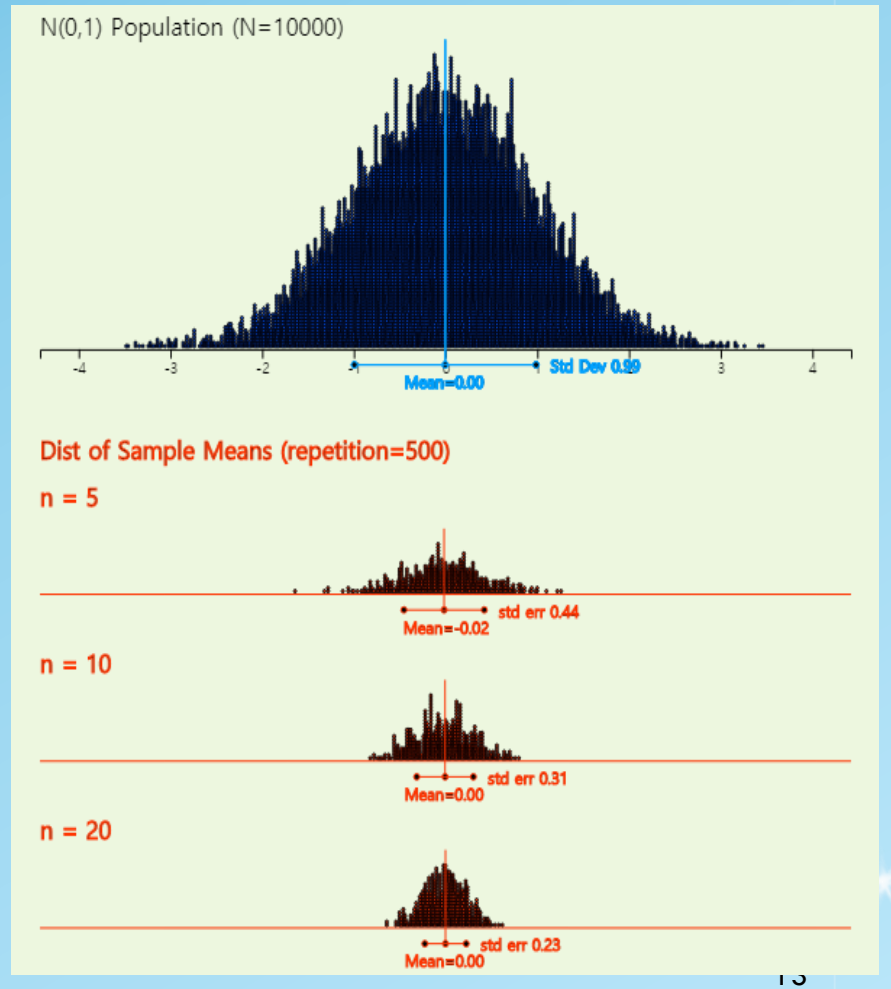
▪ Central Limit Theorem (CLT)



n1 > n2 > n3 > n4



▪ Simulation of CLT



6.2.2 Estimation of Population Mean

- When a sample survey is conducted, only one set of samples is extracted from the population to estimate the population mean.
- We consider the sample mean of the extracted sample as an estimate of the population mean. Do you think the estimation of the population mean is correct using only one set of samples?
- Sampling distribution of all possible sample means is the answer to this question. That is, whatever the population distribution is, if the sample size is large enough, all possible sample means are clustered around the population mean in the form of a normal distribution.
- Therefore, the sample mean obtained from one set of samples is usually close to the population mean, and even in the worst case, the difference from the population mean (known as an error) is not significant, so we can answer that it is possible to estimate.
- The larger the sample size, the more concentrated the sampling distribution of the sample means, which reduces this error.

6.2.2 Estimation of Population Mean

A. Point Estimation of Population Mean

- A value of the observed sample mean is called a point estimate of the population mean.
- The average value of all possible sample statistics is equal to the population parameter, and the sample statistics with this nature are called unbiased estimator.
  – sample mean is the unbiased estimator of the population mean.
- As a sample grows in size, the value of the sample statistic becomes closer and closer to the population parameter, and the sample statistic with this nature is referred to as a consistent estimator.
  – sample mean is a consistent estimator of the population mean.
- An effective estimator of the sample statistic has the least variance among several unbiased estimators of the sample statistic.
  – sample mean is an efficient estimator.

6.2.2 Estimation of Population Mean

B. Interval Estimation of Population Mean – Known Population Variance

☞ 100(1-α)% Confidence Interval for Population Mean μ

--- Population is normal and population variance σ^2 is known

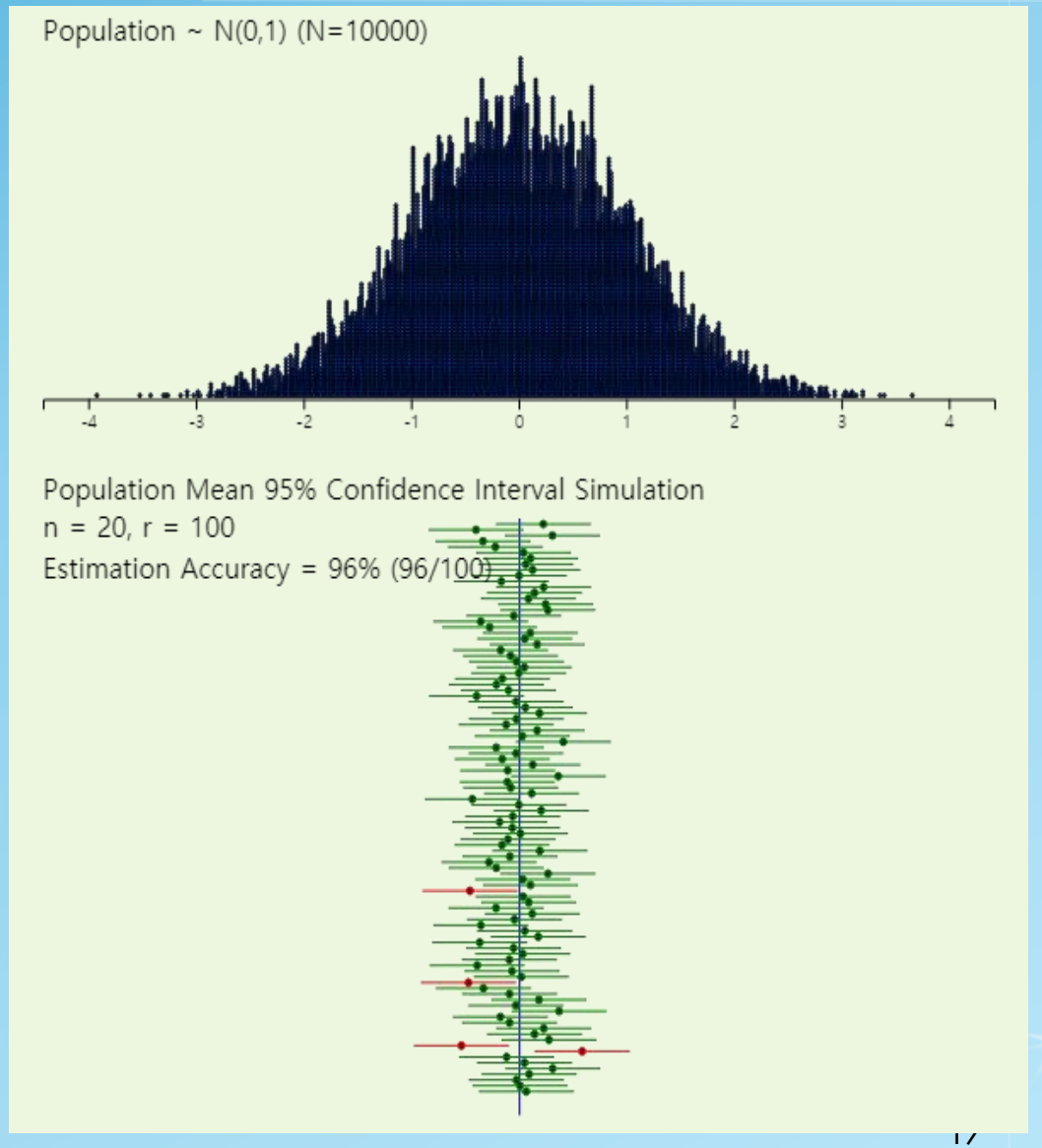
$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ , \ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- (1-α) or 100(1-α)% here is called as confidence level, which refers to the probability of intervals that will include the population mean among all intervals calculated by this interval formula. Usually, we use 0.01 or 0.05 for α. z_α is the $100(1-\alpha)$ percentile of the standard normal distribution. In other words, if Z is the random variable which follows the standard normal distribution, the probability that Z is greater than z_α is α, I.e., $P(Z > z_\alpha) = \alpha$. For example, $z_{0.025}$ = 1.96, $z_{0.95}$ = -1.645, $z_{0.005}$ = 2.575.

6.2.2 Estimation of Population Mean

B. Interval Estimation of Population Mean
- Known Population Variance

- 『eStatU』Simulation of 95% confidence interval



6.2.2 Estimation of Population Mean

[Ex 6.2.2] The average starting salary was 2.75 million won after a simple random sampling of 100 college graduates this year. Assume that the starting salary for all college graduates is a normal distribution and its standard deviation is 50,000 won.

1) Estimate the average starting salary of all college graduates.
2) Estimate a 95% confidence interval of the average starting salary of college graduates.
3) Estimate a 99% confidence interval of the average starting salary of college graduates. Compare the width of this interval to the 95% confidence interval?
4) If the sample size is 400, estimate a 95% confidence interval of the average starting salary for all college graduates. Compare the width of the interval to question 2)?

6.2.2 Estimation of Population Mean

〈Answer of Ex 6.2.2〉

1) Point estimation of the average starting salary is the sample mean which is 2.75 million won.

2) Since the 95% confidence interval implies α = 0.05, z value is as follows.

$$z_{\alpha/2} = z_{0.05/2} = z_{0.025} = 1.96$$

Therefore the 95% confidence interval is as follows.

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

$\Leftrightarrow$ (275 − 1.96(5/10), 275 + 1.96(5/10))

$\Leftrightarrow$ (274.02, 275.98)

3) Since the 99% confidence interval implies α = 0.01, z value is as follows.

$$z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.575$$

Therefore the 99% confidence interval is as follows.

(275 − 2.575(5/10), 275 + 2.575(5/10))

$\Leftrightarrow$ (273.71, 276.29)

Therefore, if the confidence level is increasing, the width of the confidence interval becomes wider.

4) If the sample size is 400, the 95% confidence interval is as follows.

(275 − 1.96(5/20), 275 + 1.96(5/20))

$\Leftrightarrow$ (274.51, 275.49)

Therefore, if the sample size is increasing, the width of the confidence interval becomes narrower which is an accurate estimation.

6.2.2 Estimation of Population Mean

B. Interval Estimation of Population Mean – Unknown Population Variance

☞ 100(1-α)% Confidence Interval for Population Mean μ

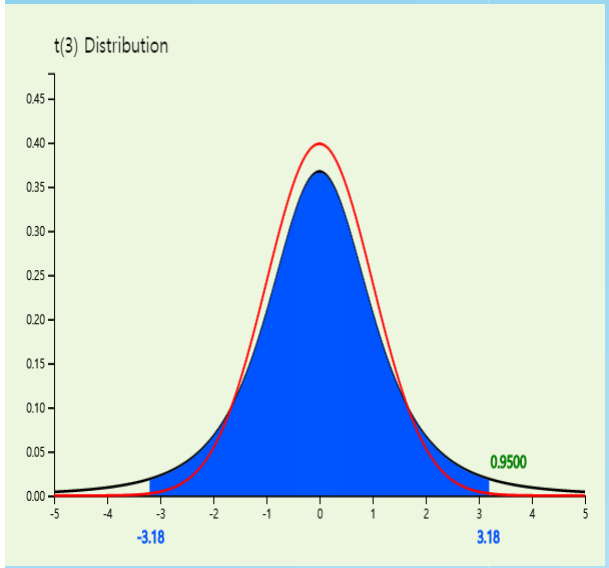
--- Population is normal and population variance σ^2 is unknown

$$\left[\overline{X} - t_{n-1;\alpha/2}\cdot\frac{S}{\sqrt{n}} \;,\; \overline{X} + t_{n-1;\alpha/2}\cdot\frac{S}{\sqrt{n}}\right]$$

n is the sample size and S is the sample standard deviation.

6.2.2 Estimation of Population Mean

- t distribution is closer to the standard normal distribution as the degree of freedom increases, usually with a degree of freedom above 30. This is why a confidence interval can be obtained approximately by using the normal distribution.
- Denote $t_{n\,;\,\alpha}$ as the $100(1-\alpha)\%$ percentile of the t distribution with degrees of freedom n. For example, $t_{7\,;\,0.05}$ is the $100(1-0.95)=95\%$ percentile of the t distribution and its value is 1.895 as <Figure 6.2.7>. In the standard normal distribution, this value was 1.645.
- Since the t distribution is symmetrical, $t_{n\,;\,\alpha} = -t_{n\,;\,1-\alpha}$. When you click on 't distribution' in the main menu of 『eStatU』 and then set the degree of freedom to 7, and the probability value in the second part below the t distribution graph to 0.95, $t_{7\,;\,0.05}$ = 1.895



6.2.2 Estimation of Population Mean

[Ex 6.2.3] Suppose we do not know the population variance In Example 6.2.2. If the sample size is 25 and the sample standard deviation is 50,000 won, estimate the mean of the starting salary of college graduates at 95% confidence level.

<Answer>

- Since we do not know the population variance, t distribution should be used for interval estimation of the population variance. Since $t_{n-1\,;\,\alpha/2} = t_{25-1\,;\,0.05/2} = t_{24\,;\,0.025} = 2.0639$, the 95% confidence interval of the population mean is as follows.

$$\left(\overline{X} - t_{n-1\,;\,\alpha/2} \cdot \frac{S}{\sqrt{n}} \;,\; \overline{X} + t_{n-1\,;\,\alpha/2} \cdot \frac{S}{\sqrt{n}} \right)$$

$\Leftrightarrow$ (275 − 2.0639(5/5), 275 + 2.0639(5/5))

$\Leftrightarrow$ (272.9361, 277.0639)

Note that the smaller the sample size, the wider the interval width.

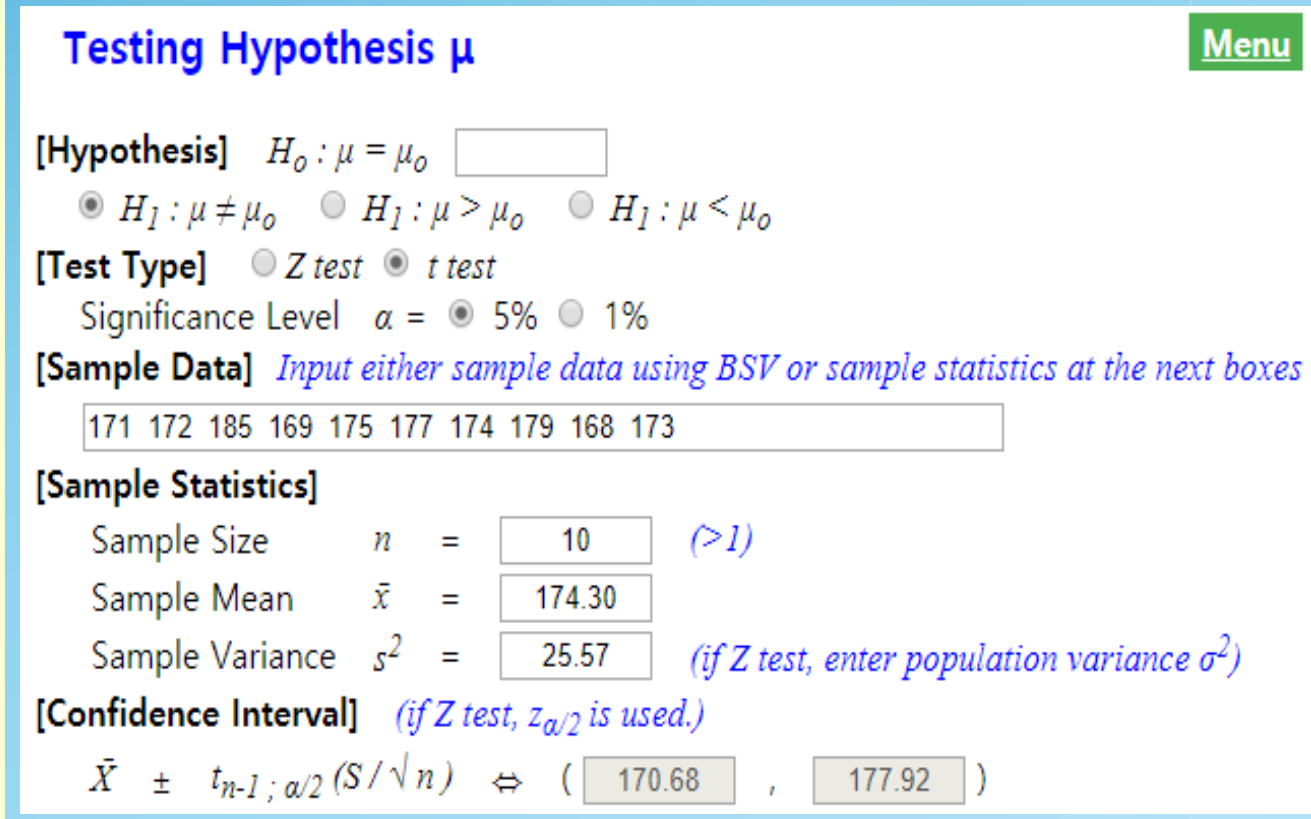
6.2.2 Estimation of Population Mean

[Example 6.2.4] The following data shows a simple random sampling of 10 new male students this year to investigate the heights of university students. Use 『eStatU』 to make a 95% confidence interval of the height of college freshmen.

171 172 185 169 175 177 174 179 168 173

〈Answer〉

- Click 'Testing Hypothesis μ' at the 『eStatU』 and enter data at the [Sample Data] box.
- Then the confidence intervals [170.68, 177.92] are calculated using the t distribution.



6.3 Sampling Distribution of Sample Variances and Estimation of Population Variance

6.3.1 Sampling Distribution of Sample Variances

[Ex 6.3.1] Let's consider the data again in Example 6.2.1 which is the number of years of service for the five salespeople.

6, 2, 4, 8, 10

1) Calculate the population variance.
2) Find all possible samples of size 2 with replacement and calculate the sample variance of each sample. In addition, calculate the average and variance of all of these sample variances and compare them to the population variance.
3) Find a frequency distribution of all possible sample variances and draw a bar chart.

〈풀이〉

1) The population mean is $\mu = 6$ and variance is 8.

6.3.1 Sampling Distribution of Sample Variances

⟨Answer of Ex 6.2.1⟩

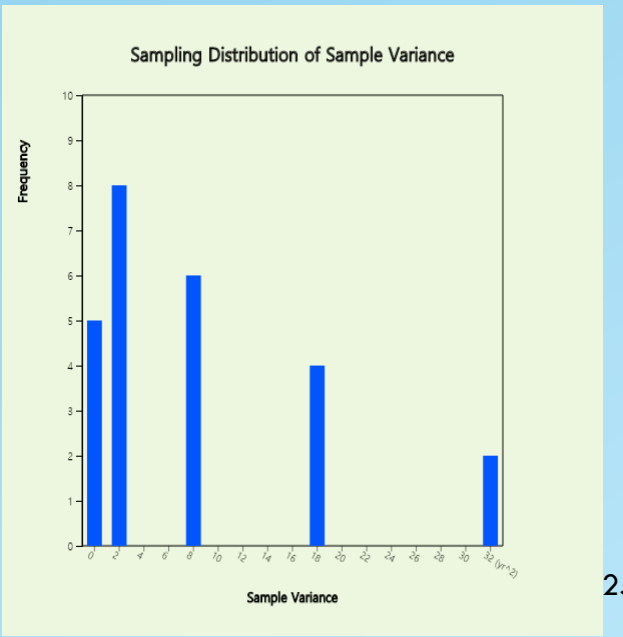
2) All possible samples of size 2 with replacement from the population and the sample variance of each sample

Sample	s^2	Sample	s^2	Sample	s^2	Sample	s^2	Sample	s^2
2,2	0	4,2	2	6,2	8	8,2	18	10,2	32
2,4	2	4,4	0	6,4	2	8,4	8	10,4	18
2,6	8	4,6	2	6,6	0	8,6	2	10,6	8
2,8	18	4,8	8	6,8	2	8,8	0	10,8	2
2,10	32	4,10	18	6,10	8	8,10	2	10,10	0

- the average of all possible sample variances is the same as the population variance which means the sample variance is the unbiased estimate of the population variance.

- Sampling distribution of sample variace

Sample variance	Frequency	Relative frequency
0	5	0.20
2	8	0.32
8	6	0.24
18	4	0.16
32	2	0.08
	25	1.00



6.3.1 Sampling Distribution of Sample Variances

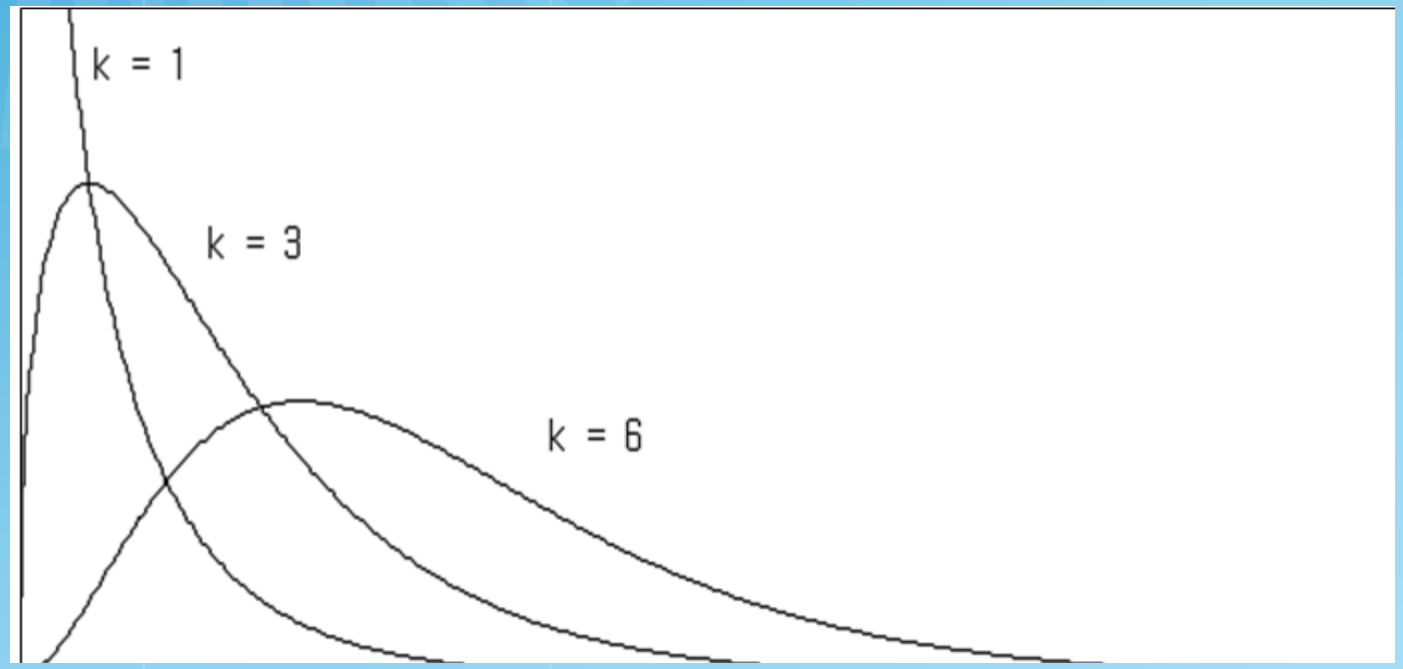
☞ The sampling distribution of the sample variance

When the population is normally distributed and a sample of size n is selected randomly with replacement, the distribution of sample variance multiplied by the specific constant follow the chi-square distribution with n- degree of freedom as follows

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

6.3.1 Sampling Distribution of Sample Variances

- This chi-square distribution is a family of distributions depending on a degree of freedom, such as χ_1^2, χ_2^2, ... , χ_{27}^2, ... etc. The chi-square distribution is an asymmetrical distribution as <Figure 6.3.2>. If the degree of freedom is small, the shape of the chi-square distribution is much skewed to the right.



6.3.2 Estimation of Population Variance

- Examples of estimating population variance.
  - Two companies supply bolts to an automaker. Bolts are defective even if they are too large or too small in diameter. The automaker wants to recognize the variance of bolt diameters supplied by each bolt company and use them as data for decision making.
  - To evaluate the difficulty of the college entrance exam conducted this year, the variance of the exam scores is calculated and compared with the variance of the exam score of previous year.

☞ Point estimation of population variance σ^2

- Sample variance S^2 (S^2 is an unbiased estimatior of σ^2)

☞ Point estimation of population standard deviation σ

- Sample variance S (S is not an unbiased estimatro of σ)

6.3.2 Estimation of Population Variance

☞ 100(1-α)% Confidence interval of the population variance σ^2
- when population is normally distributed

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} , \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}} \right]$$

☞ 100(1-α)% Confidence interval of the population standard deviation σ
- when population is normally distributed and sample is large

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}} , \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}} \right]$$

S^2 is the sample variance, $\chi^2_{k;p}$ is the upper 100(1-p) percentile of the chi-square distribution with k degree of freedom.

6.3.2 Estimation of Population Variance

[Ex 6.3.2] A survey for the starting salary of 25 college graduates this year shows the sample standard deviation is 5 (1000$). Find point estimation and 95% confidence interval of the population variance and the population standard deviation. Assume that the population is normally distributed.

<Answer>

- The point estimate of the population variance for the starting salary of college graduate is the sample variance, so $s^2 = 5^2 = 25$. Since the point estimate of the population standard deviation is the sample standard deviation, so $s = 5$.
- The 95% confidence interval of the population variance is as follows.

$$\left[\frac{(n-1)S^2}{\chi^2_{25-1;\,0.05/2}} , \frac{(n-1)S^2}{\chi^2_{25-1;\,1-0.05/2}} \right]$$

$$\Leftrightarrow \left[\frac{(25-1)5^2}{39.364} , \frac{(25-1)5^2}{12.401} \right]$$

$$\Leftrightarrow [\ 15.242\ ,\ 48.383\ ]$$

- The 95% confidence interval of the population standard deviation is as follows.

$$[\ \sqrt{15.242}\ , \sqrt{48.383}\ ] \Leftrightarrow [3.904\ ,\ 6.956]$$

6.3.2 Estimation of Population Variance

[Ex 6.3.3] The height data of 10 male freshman samples is as follows.

171 172 185 169 175
177 174 179 168 173

Use 『eStatU』 to make a 95% interval estimate of the population variance.

Testing Hypothesis σ^2 Menu

[Hypothesis] $H_0 : \sigma^2 = \sigma_0^2$ ( ) (> 0)

$H_1 : \sigma^2 \neq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$

[Test Type] χ^2 *test*

Significance Level α = 5% 1%

[Sample Data] *Input either sample data using BSV or sample statistics at the next boxes*

171 172 185 169 175 177 174 179 168 173

[Sample Statistics]

Sample Size n = 10 (> 1)

Sample Variance s^2 = 25.567 (> 0)

[Confidence Interval]

$\left((n-1)S^2 / \chi^2_{n-1\,;\,\alpha/2}\,,\ (n-1)S^2 / \chi^2_{n-1\,;\,1-\alpha/2} \right) \Leftrightarrow$ (12.10 , 85.21)

6.4 Sampling Distribution of Sample Proportions and Estimation of Population Proportion

6.4.1 Sampling Distribution of Sample Proportions

[Example 6.4.1] Let's call 10 employees of a company a population. When the employees' satisfaction level with the company is investigated and the satisfaction is expressed as 1, the complaint is 0 as follows.

1 0 1 1 0 1 1 0 0 1

That is, the population proportion p of the satisfaction is 0.6.
Consider all possible samples of size 5 with replacement to obtain a sampling distribution of sample proportions.

[Answer of Ex 6.4.1]

Table 6.4.1 All possible sample cases

Sample Case		Number of cases	
all unsatisfactory	(0,0,0,0,0)	${}_5C_0 \times 4 \times 4 \times 4 \times 4 \times 4$ =	1024
1 satisfactory	(0,0,0,0,1)	${}_5C_1 \times 4 \times 4 \times 4 \times 4 \times 6$ =	7680
2 satisfactory	(0,0,0,1,1)	${}_5C_2 \times 4 \times 4 \times 4 \times 6 \times 6$ =	23040
3 satisfactory	(0,0,1,1,1)	${}_5C_3 \times 4 \times 4 \times 6 \times 6 \times 6$ =	34560
4 satisfactory	(0,1,1,1,1)	${}_5C_4 \times 4 \times 6 \times 6 \times 6 \times 6$ =	25920
5 satisfactory	(1,1,1,1,1)	${}_5C_5 \times 6 \times 6 \times 6 \times 6 \times 6$ =	7776
		계	100000

Table 6.4.2. Sampling Distribution of Sample Proportions

Sample case	$\hat{p}$	Frequency	Relative Frequency
all unsatisfactory	0.0	1024	0.01024
1 satisfactory	0.2	7680	0.07680
2 satisfactory	0.4	23040	0.23040
3 satisfactory	0.6	34560	0.34560
4 satisfactory	0.8	25920	0.25920
5 satisfactory	1.0	7776	0.07776
	Total	100000	1.0

6.4.1 Sampling Distribution of Sample Proportions

☞ Sampling distribution of sample proportions --- infinite population

Assume the population proportions is p. If the sample size is large, the sampling distribution of the sample proportion $\hat{p}$ is approximately a normal distribution with mean p and variance $p(1-p)/n$.

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

6.4.1 Sampling Distribution of Sample Proportions

[Ex 6.4.2] Let's say 3% of semiconductors made in a semiconductor factory are defective. When 300 samples were taken without replacement, the sample proportion for defective products was 2%. Find out where this sample rate is located among all possible sample proportions. What is the probability that the sample proportion is greater than 2%?

<Answer>

- Since the sampling distribution of the sample proportions is approximately normal distribution, $\hat{p} \sim N\left(0.03, \frac{0.03(1-0.03)}{300}\right)$, the probability can be calculated as follows.

$$P(\hat{p} > 0.02) = P(Z > (0.02-0.03)/0.00985)$$
$$= P(Z > -1.02)$$
$$= 1 - P(Z \leq -1.02)$$
$$= 1 - 0.1539 = 0.8461$$

6.4.2 Estimation of Population Proportion

- Some practical examples to estimate the proportion of unknown populations.

- What is the approval rating of a particular political party in this year's election?
- What percentage of the nation's current unemployment rate is?
- What percentage of defective products do we have here when we import 10,000 car accessories?

☞ Point estimation of population proportion p is the sample proportion $\hat{p}$.

Sample proportion ($\hat{p}$) is an unbised, efficient and consistent estimator of the population proportion p and the estimate of the standard error of $\hat{p}$ is $\sqrt{\hat{p}(1-\hat{p})/n}$

6.4.2 Estimation of Population Proportion

☞ Interval estimation of the population proportion p with 100(1-α)% confidence interval – large sample case

$$\left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \, , \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Criteria of large sample size n is as follows.

$$n\hat{p} > 5, \quad n(1-\hat{p}) > 5$$

6.4.2 Estimation of Population Proportion

[Example 6.4.3] A student running for president of a university had a simple survey of 200 students to find out his approval ratings, and found that 120 students supported him. Estimate the population's approval rating, and find a 95% confidence interval. Check the interval estimation using 『eStatU』

<Answer>

- Estimation of the population approval rating is the sample proportion.

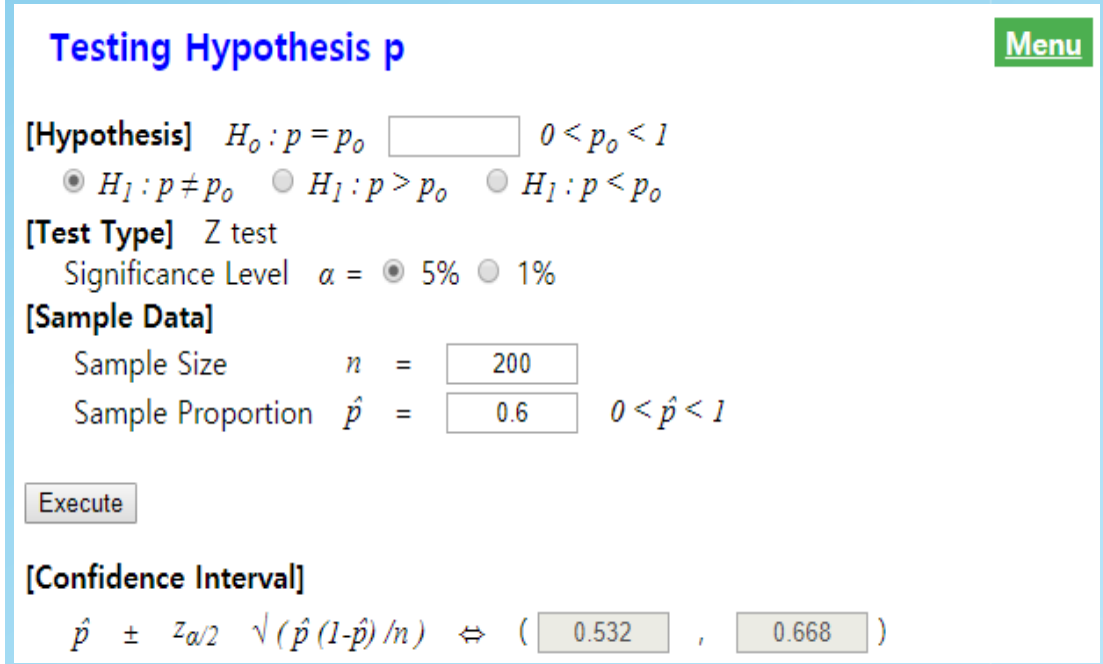
$$\hat{p} = \frac{120}{200} = 0.6$$

- The 95% confidence interval is as follows.

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$\Leftrightarrow \left[0.6 - 1.96\sqrt{\frac{0.6(1-0.6)}{200}},\ 0.6 + 1.96\sqrt{\frac{0.6(1-0.6)}{200}}\right]$$

$$\Leftrightarrow [0.532,\ 0.668]$$



6.5 Determination of Sample Size

6.5.1 Determination of Sample Size to Estimate Population Mean

☞ Determination of sample size to estimate the population mean

$$n = \lceil \frac{z_{\alpha/2}\,\sigma}{d} \rceil^2$$

- Since the population standard deviation σ in the above equation is unknown usually, the estimated value from past experience or data obtained from preliminary survey is used. The estimation of the population standard deviation σ through preliminary survey is done by using the range as follows.

$$\hat{\sigma} = \frac{range}{4} = \frac{\max - \min}{4}$$

6.5.1 Determination of Sample Size to Estimate Population Mean

[Ex 6.5.1] The standard deviation of the life of a light bulb produced at a plant is usually 100 hours. To estimate the average life of a bulb at a 95% confidence level, calculate the size of the sample to be within 20 hours of the error.

<Answer>

- $n = \lceil \frac{z_{\alpha/2}\,\sigma}{d} \rceil^2 = (\frac{1.96 \times 100}{20})^2 = 9.8^2 = 96.04$

  Hence the sample size is 97 approximately.

6.5.1 Determination of Sample Size to Estimate Population Proportion

☞ Determination of sample size to estimate the population proportion

$$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{d}\right)^2$$

- In the above equation, the value of the sample proportion $\hat{p}$ is usually used by the value of past experience or preliminary investigation. However, if there is no information about the population proportion, the value of 0.5 is usually used.

6.5.1 Determination of Sample Size to Estimate Population Proportion

[Ex 6.5.2] For this year's presidential election, a survey for estimating the 95% confidence interval of candidate's approval rating is conducted. Obtain the size of the sample to be within 2.5% of the error bound.

<Answer>

- Since we do not have information on the population proportion, assume $\hat{p}$ = 0.5. Then the sample size n can be calculated as follows.

$$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{d}\right)^2$$

$$n = 0.5(1-0.5)\frac{1.96^2}{0.025^2} = 1536.6$$

- Therefore, samples must be extracted with at least 1537 persons to limit the error bound to 2.5%. Various opinion polls conducted in Korea often show that the size of the sample is around 1,500 which implies that the margin of error does not exceed 3 percent.

6.6 Summary

- Simple Random Sampling
- Sampling Distribution of Sample Means
  - Central Limit Theorem
  - Estimation of Population Mean : Z or t distribution
- Sampling Distribution of Sample Variances
  - Estimation of Population Variance : ChiSquare distribution
- Sampling Distribution of Sample Proportions
  - Estimation of Population Proportions : Normal distribution
- Determination of Sample Size
  - To estimate population mean and population proportion

Thank you