Introduction to Statistics and Data Science using *eStat*

Chapter 6 Sampling Distribution and Estimation

6.2 Sampling Distribution of Sample Means and Estimation of Population Mean

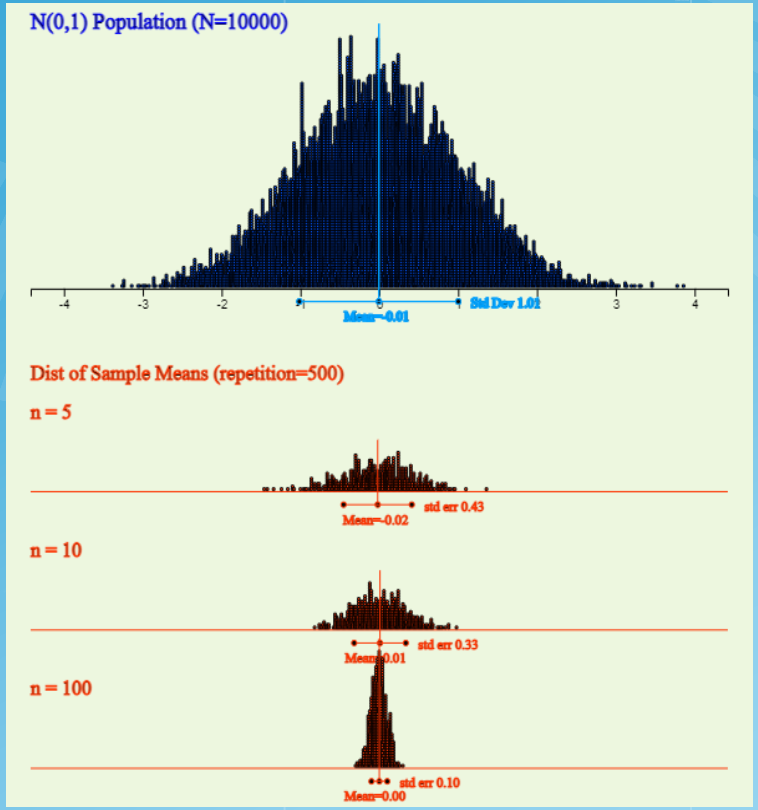
Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

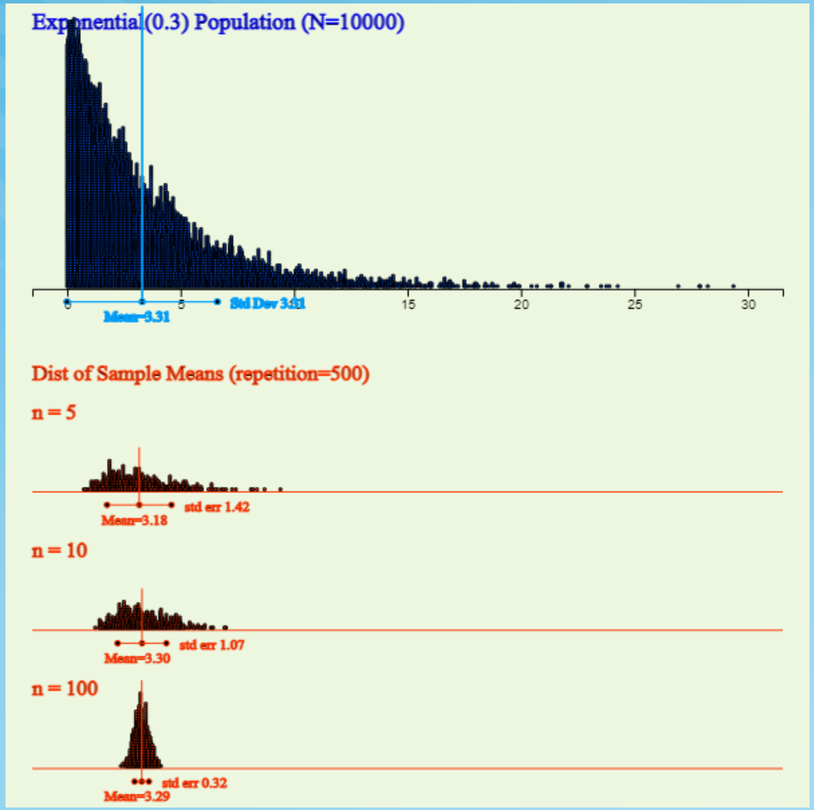
6.2.2 Estimation of Population Mean

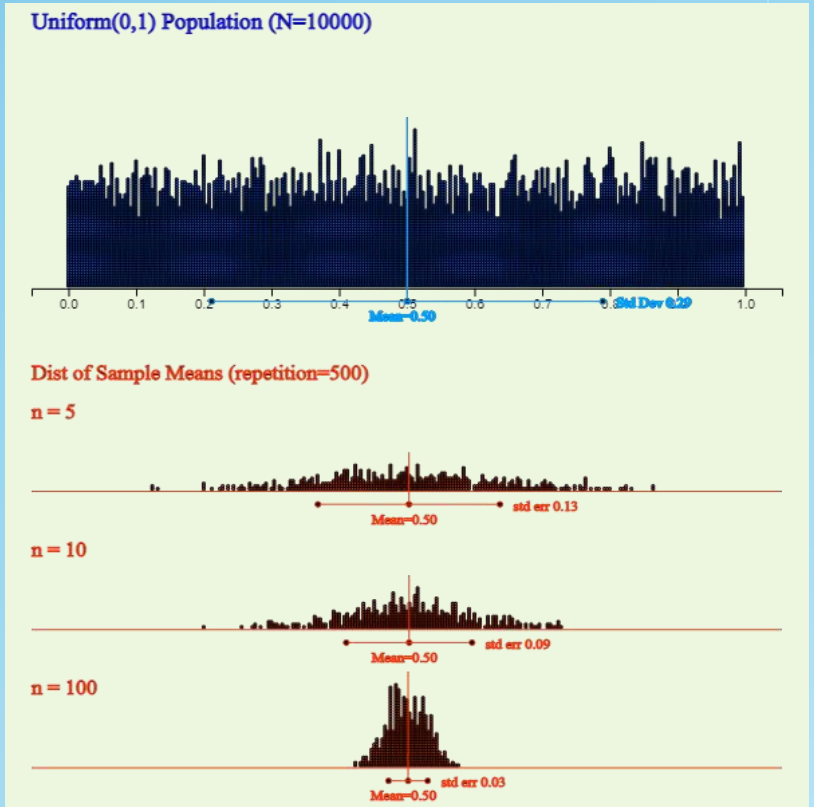
- Sample survey extracts only one set of samples from population.
  => sample mean of samples is an estimate of the population mean.
  => Is this sample mean is a good estimate for population mean?
  just only one set of samples?

- Sampling distribution of all possible sample means is the answer
  => if sample size is large enough, all possible sample means are clustered around the population mean in the form of a normal distribution.
  => Most of sample mean obtained from one set of samples is close to the population mean.
  => Even in worst case, error from population mean is not so significant.
  => The larger the sample size, the more concentrated the sampling distribution of the sample means, which reduces this error.

▪ Central Limit Theorem(CLT) - Simulation







6.2.2 Estimation of Population Mean

A. Point Estimation of Population Mean

- A value of observed sample mean is a point estimate of population mean.
- If the average of all possible sample statistics is equal to the population parameter, the sample statistic is called an unbiased estimator.
  - sample mean is an unbiased estimator of the population mean.
- When sample size grows, if sample statistic becomes closer to the population parameter, the sample statistic is called a consistent estimator.
  - sample mean is a consistent estimator of the population mean.
- If sample statistic has the least variance among several unbiased estimators, the sample statistic is called an effective estimator.
  - sample mean is an efficient estimator of the population mean.

6.2.2 Estimation of Population Mean

B. Interval Estimation of Population Mean – Known Population Variance

- 100(1-α)% Confidence Interval for Population Mean
  - Assume a population is a normal distribution and σ^2 is known.

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

- usually, $\alpha = 0.01\ or\ 0.05$ which is margin of error
  100(1- α)% is called a confidence level,

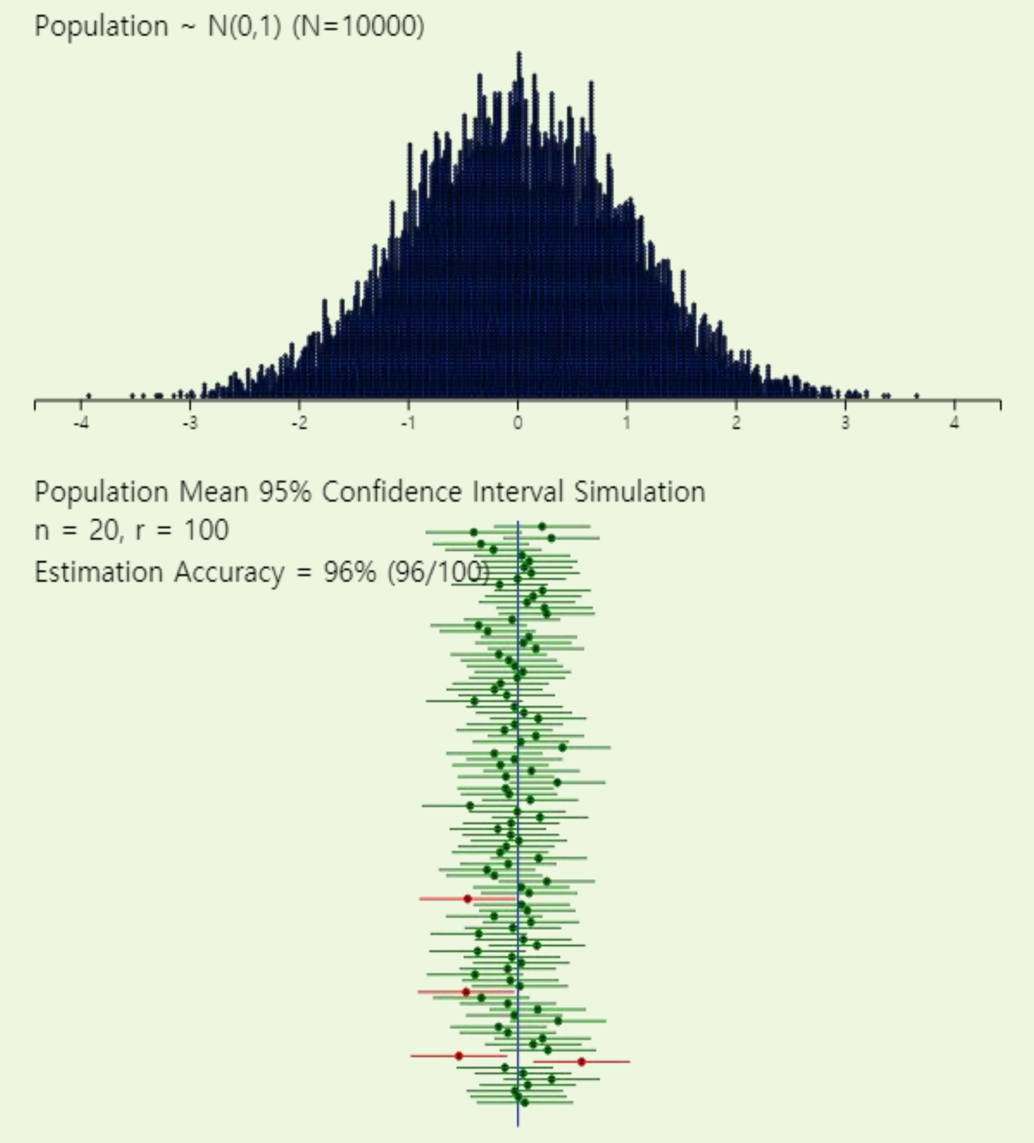
- 95% confidence interval $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$, $z_{0.05/2} = z_{0.025} = 1.96$
- 99% confidence interval $\bar{X} \pm 2.575\frac{\sigma}{\sqrt{n}}$, $z_{0.01/2} = z_{0.005} = 2.575$

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 275 (unit man 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 5 (unit man 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
  - Assume a population is a normal distribution and σ^2 is known.
  - S is sample standard deviation

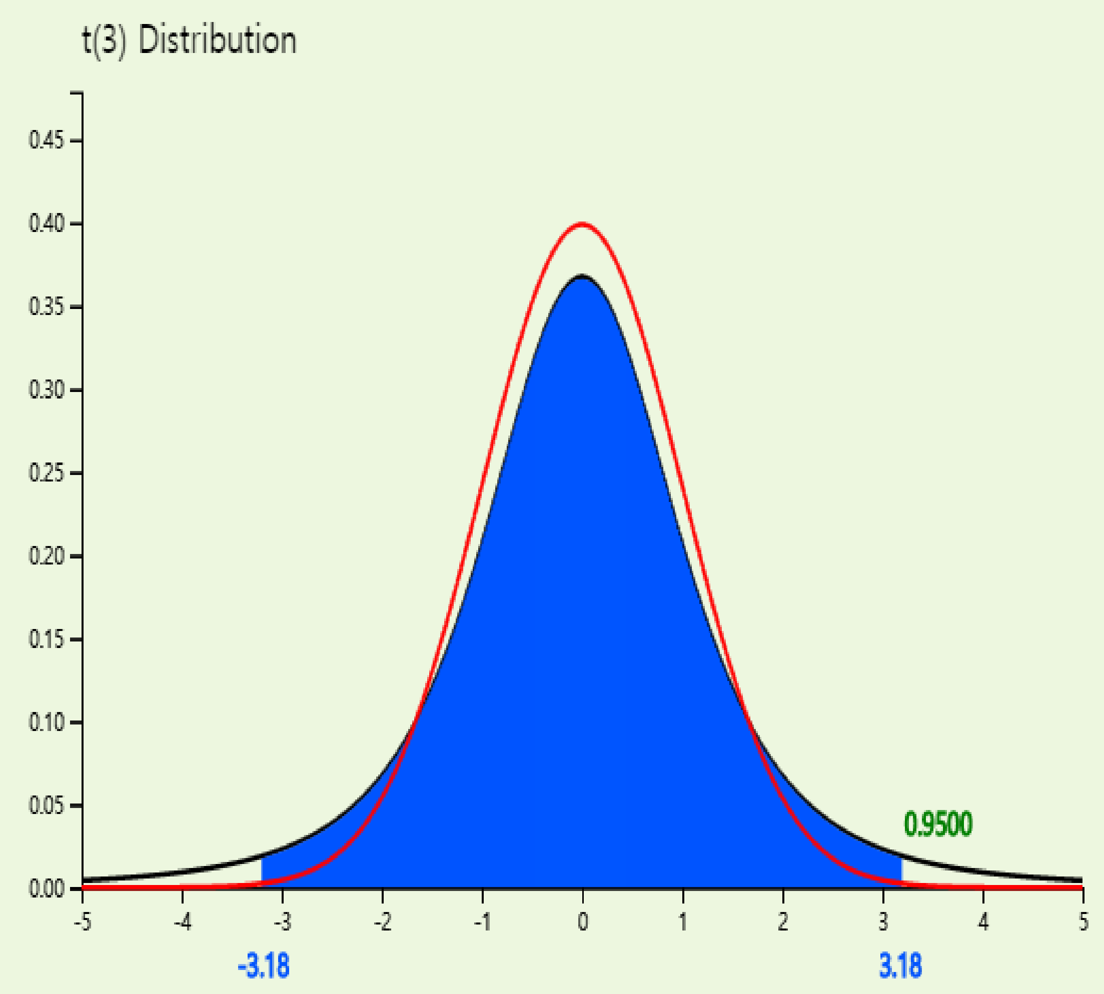
$$\left[\bar{X} - t_{n-1;\,\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + t_{n-1;\,\alpha/2}\frac{S}{\sqrt{n}}\right]$$

(

- usually $\alpha = 0.01\ or\ 0.05$ which is margin of error
  100(1- α)% is called a confidence level,

6.2.2 Estimation of Population Mean

- t distribution was studied by a statistician W. S. Gosset in Ireland.
- t distribution is not just a single distribution, but it is a family of distributions under a parameter called a degree of freedom, 1,2, … , 30, … and denoted as $t_1, t_2, \dots, t_{30}, \dots$
- The shape of the t distribution is symmetrical about zero (y axis), similar to the standard normal distribution, but it has a tail that is flat and longer than the standard normal distribution.



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 5 (unit man 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 mean.
- 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:

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

⇔ (275 - 2.0639 (5/5) , 275 + 2.0639 (5/5))

⇔ (272.9361, 277.0639)

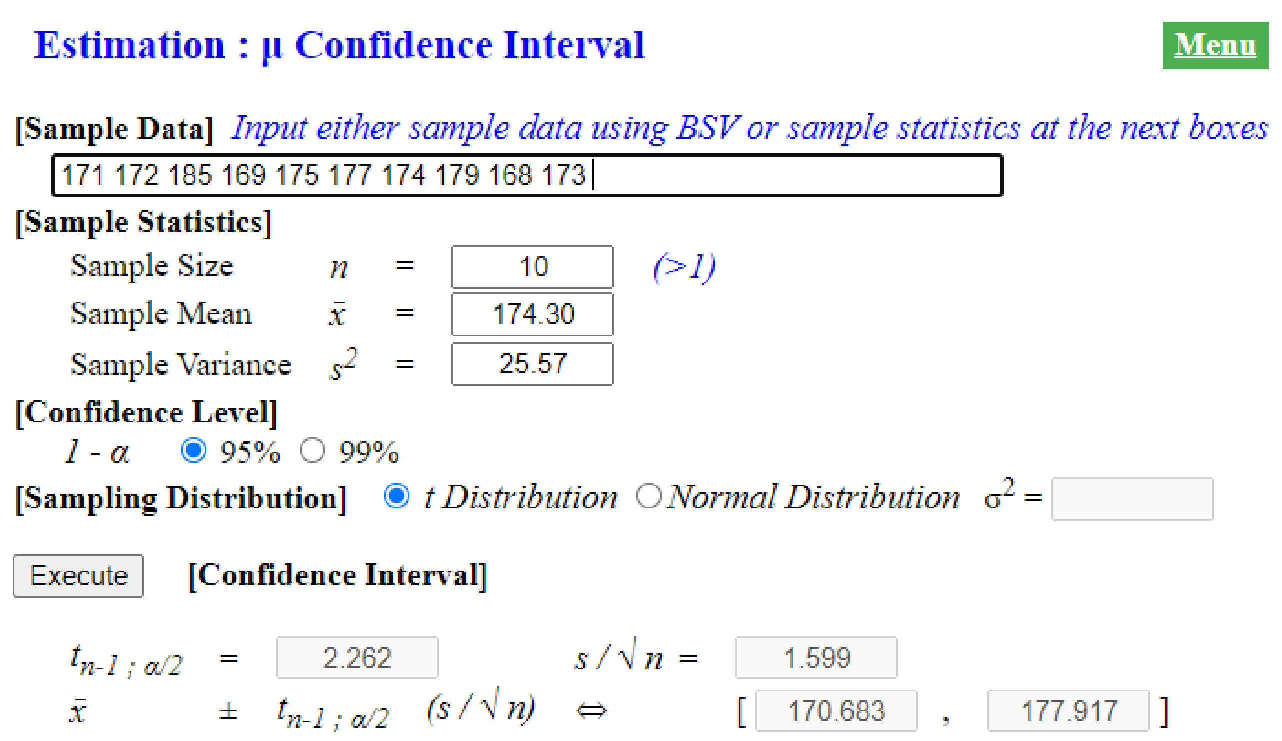
6.2.2 Estimation of Population Mean

[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 ‘Estimation: μ Confidence Interval’ 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.



Thank you