LECTURE 9
INFERENTIAL STATISTICS: ESTIMATION

PSY2002

Hye Won Suk

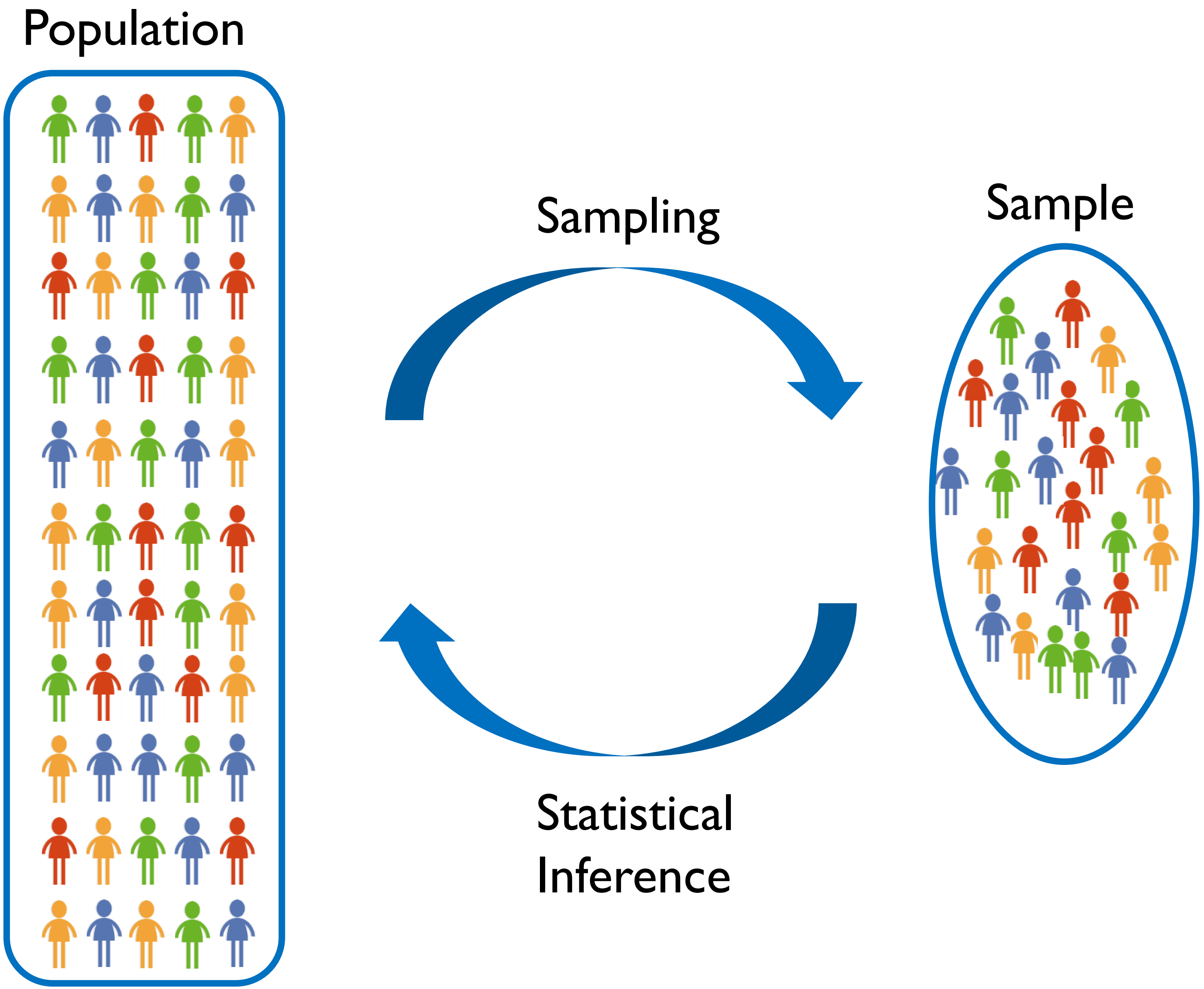
STATISTICAL INFERENCE

- When discussing the sampling distribution of mean, we assumed that population parameters, e.g., the population mean (μ) and population standard deviation (σ) are known.

- In many cases, however, we don't have population information. That is, we don't know population parameters such as population mean and population standard deviation.

- Rather, what we know is sample information. We can obtain sample statistics such as sample mean and sample standard deviation from the sample that we observe.

- <u>Statistical inference</u> is the process of drawing conclusions about the population based on information in a sample.


Population
Sampling
Sample
Statistical
Inference

STATISTICAL INFERENCE

- We will learn the following statistical inference procedures.

- Estimation
    - Point estimation
    - Interval estimation

- Hypothesis Testing
    - Z-test
    - t-test
    - Analysis of Variance (ANOVA)
    - And so on…

ESTIMATION

- A population parameter (모수치) is a quantity that describes a characteristic of a population.

- A sample statistic (통계치, 통계량) is a quantity that describes a characteristic of a sample.

- <u>Estimation</u> (추정) is the process of using known sample statistics to make inference about population parameters. Estimation enables us to make educated guesses about the value of unknown population parameters.

EXAMPLE

- A professor at Sogang University developed an on-line statistics course and she wants to estimate the population mean and variance of the satisfaction scores of the students for this course. Last semester, 100 students were enrolled in her class. The mean satisfaction score of the 100 students was 6.7 (out of 10) and the variance (obtained using n-1 formula) was 2.1.

- What is the estimate of the population mean satisfaction level of her course?

- What is the estimate of the population variance of the satisfaction score?

POINT ESTIMATION

- Point estimation (점추정) is to find a single value that represents our best guess of the population parameter.

- An unbiased estimator can be used as a point estimator.
  - Sample mean $\left(\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}\right)$ is the unbiased estimator of the population mean.
  - $E(\bar{X}) = \mu$

  - Sample variance $\left(s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}\right)$ is the unbiased estimator of the population variance.
  - $E(s^2) = \sigma^2$

- The sample mean satisfaction score is 6.7 ($\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 6.7$).
- We can estimate that the population mean satisfaction level is 6.7.

- The sample variance of satisfaction score is 2.1 ($s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = 2.1$).
- We can estimate that the population variance of satisfaction score is 2.1.

- A point estimate will not match the population parameter exactly. However, it is an unbiased and educated guess, given the sample.

INTERVAL ESTIMATION

- In the previous example, the sample mean (obtained from the given sample) was used as a point estimator of the population mean.

- However, the sample mean varies from sample to sample. If the researcher has had a different sample of 100 students, the sample mean would have not been 6.7, but a different value.

- Even though the sample mean that she obtained is the most plausible value for the population, there is some uncertainty.

- To take into account this uncertainty in our estimation (sample-to-sample variation), we can use an interval estimation.

- Interval estimation (구간추정) gives a range of plausible values, or an interval, for a population parameter rather than a single value.

- The common form for an interval estimate is

$$\boxed{Point\ estimate\ \pm margin\ of\ error}$$

- For example, an interval estimate for the population mean is given by the form

$$\boxed{\bar{X} \pm margin\ of\ error}$$

MARGIN OF ERROR

- How do we obtain the margin of error (오차범위)?

- There are two things to consider when we determine the margin of error.

  - How much uncertainty we can tolerate, i.e., how much confidence we want.

  - How much the sample statistic (or point estimate) varies from sample to sample.

1. CONFIDENCE LEVEL

- How much uncertainty we can tolerate:

  - If the margin of error is very large, the interval is very wide. In that case, we are almost 100% sure that the population parameter lies within the interval.

  - On the contrary, if the margin of error is very small, the interval is very narrow. In that case, we are less sure that the population parameter lies within that interval.

- In the previous example, the satisfaction score is out of 10. Let's consider the following three statements:

"I estimate that the population mean satisfaction score lies between 0 and 10."

- Of course, it is correct. We are 100% sure that it is correct.

"I estimate that the population mean satisfaction score lies between 1 and 9."

- Not 100% sure, but it is very highly likely that it is correct.

"I estimate that the population mean satisfaction score lies between 4 and 6."

- Could be, but not so sure if it is correct.

- We can obtain an interval estimate, or range of plausible values, of a population parameter with certain degree of confidence.

- For example,
  - 100% confidence interval
  - 95% confidence interval
  - 75% confidence interval

- A higher confidence level implies a wider confidence interval.

- In statistics, 95% and 99% confidence intervals are commonly used.

2. STANDARD ERROR

- The margin of error also depends on how much the sample statistic (or point estimate) varies from sample to sample:
  - If the sample statistic widely varies from sample to sample, the sample statistic is not precise as a point estimate for the parameter. Therefore, the margin of error should be large to obtain a precise interval estimate.
  - On the contrary, if the sample statistic does not vary much from sample to sample, the sample statistic is a precise point estimate for the parameter. Therefore, the margin of error does not need to be large to obtain a precise interval estimate.

MARGIN OF ERROR

- In sum, the margin of error can be obtained by multiplying two quantities.

$$\boxed{\text{Margin of error} = (\text{Critical value}) \times (\text{Standard error})}$$

  - The critical value reflects the confidence level.
  - The standard error reflects the variability (standard deviation) of the sample statistic.

EXAMPLE

- A professor at Sogang University developed an on-line statistics course and she wants to estimate the population mean satisfaction level for this course. Last semester, 100 students were enrolled in her class. The mean satisfaction score of the 100 students was 6.7 (out of 10). Let's assume that the population variance is known to be 2.4.

  - $\bar{X} = 6.7$
  - $\sigma^2 = 2.4$
  - $n = 100$

- Find the 95% confidence interval of the population mean satisfaction level for her course.

SAMPLING DISTRIBUTION

- To obtain the margin of error, we can use the sampling distribution of mean.
- By the central limit theorem, we know that
  - The sampling distribution of mean is a normal distribution (because the sample size $n = 100$ is greater than 30).
  - The expected value of the sample mean :

$$\mu_{\bar{X}} = \mu$$

  - The standard error of the sample mean :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- That is, the sample means are normally distributed with the mean of μ and standard deviation of $\frac{\sigma}{\sqrt{n}}$.

- From the standard normal distribution, we know that
  - $P(-1.96 < Z < 1.96) = 0.95$
  - $P(-1.96 < Z < 1.96)$

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

$$= P\left(-1.96 < \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} < 1.96\right)$$

$$= P(-1.96\sigma_{\bar{X}} < \bar{X} - \mu_{\bar{X}} < 1.96\sigma_{\bar{X}})$$

$$= P(-\bar{X} - 1.96\sigma_{\bar{X}} < -\mu_{\bar{X}} < -\bar{X} + 1.96\sigma_{\bar{X}})$$

$$= P(\bar{X} + 1.96\sigma_{\bar{X}} > \mu_{\bar{X}} > \bar{X} - 1.96\sigma_{\bar{X}})$$

$$= P(\bar{X} - 1.96\sigma_{\bar{X}} < \mu_{\bar{X}} < \bar{X} + 1.96\sigma_{\bar{X}})$$

$$\mu_{\bar{X}} = \mu$$
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- $P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$

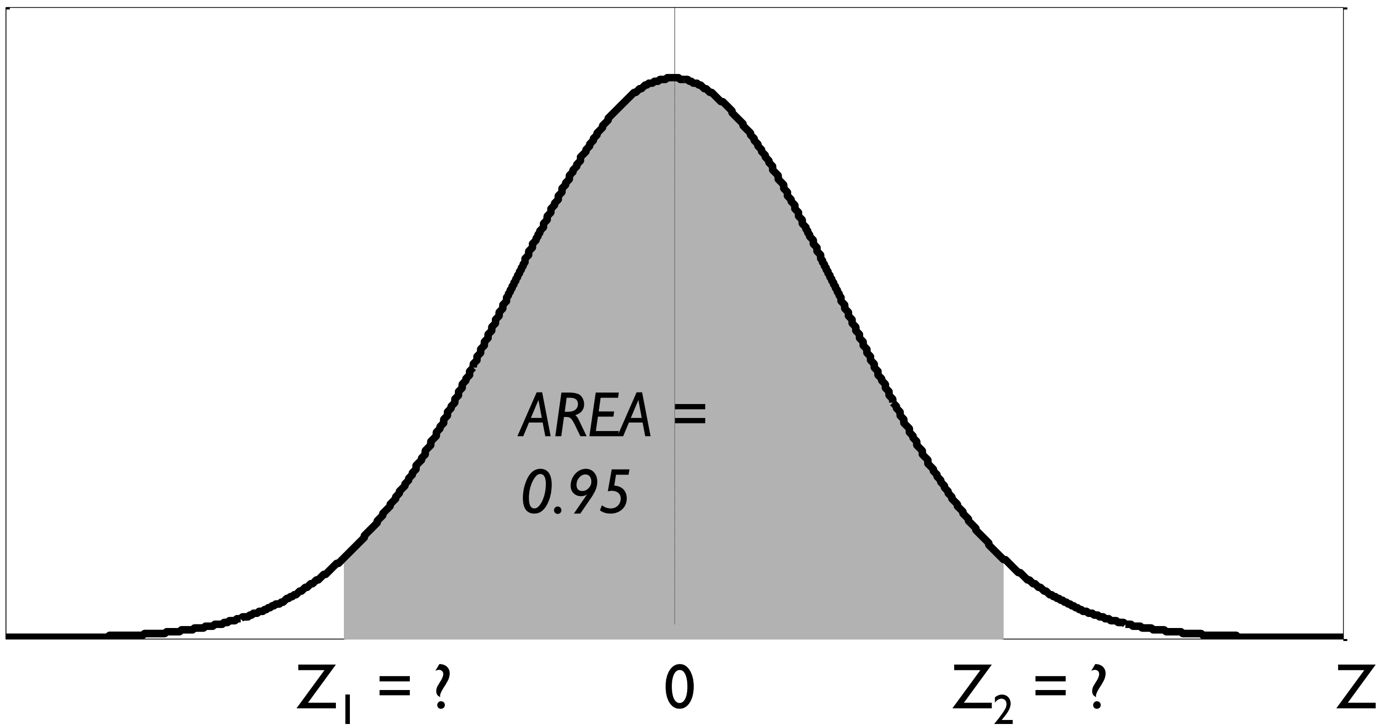
- $1.96\frac{\sigma}{\sqrt{n}}$ indicates the margin of error in estimating the population mean using the sample mean at the confidence level of 95%.

Margin of error = (Critical value) x (Standard error)

- The critical value = 1.96
- The standard error = $\frac{\sigma}{\sqrt{n}}$

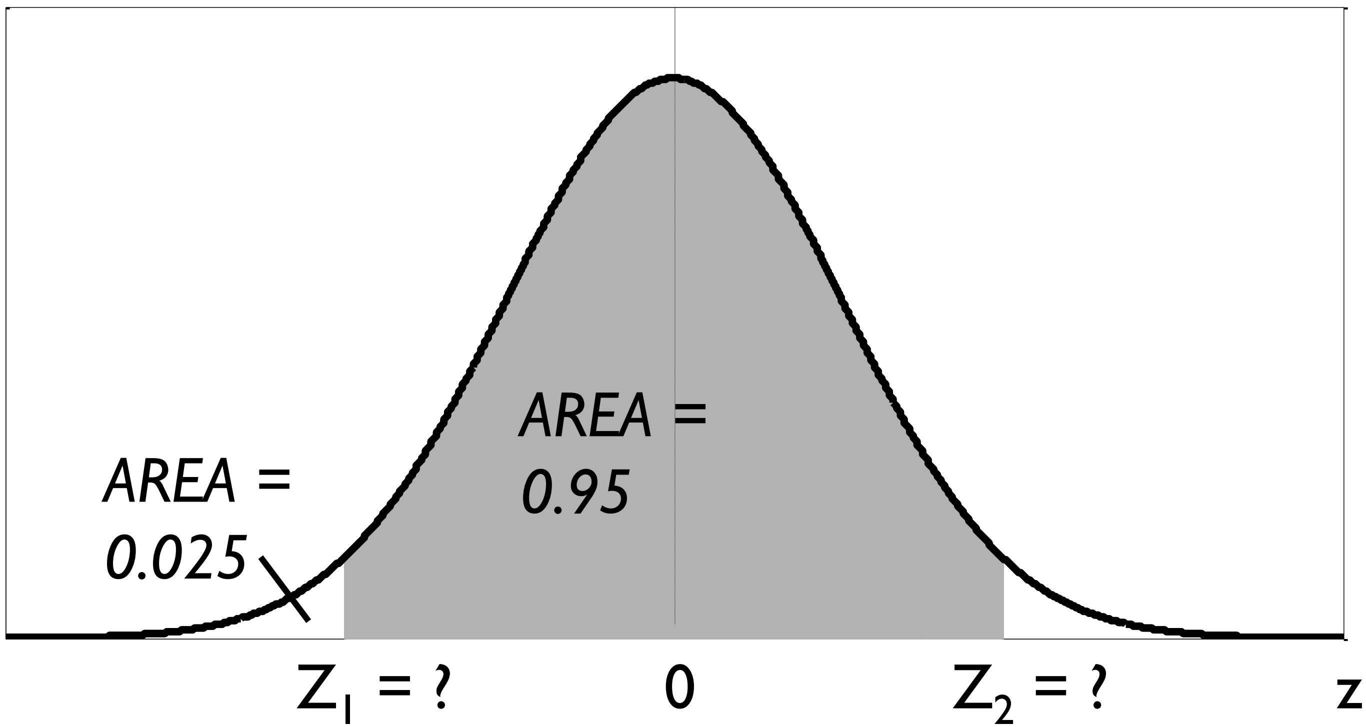
EXAMPLE

- Step 1: Find the positive critical value for the confidence level of 95%.



EXAMPLE

- Step 1: Find the positive critical value for the confidence level of 95%.



$P(Z < Z_1) = 0.025$, $Z_1 = -1.96$

$Z_2 = 1.96$

EXAMPLE

- Step 2: Find the standard error.
  - Standard error of mean ($\sigma_{\bar{X}}$)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2.4}}{\sqrt{100}} = \sqrt{\frac{2.4}{100}} = \sqrt{0.024} = 0.15$$

EXAMPLE

- Step 3: Compute the margin of error.

$$\begin{aligned}\text{Margin of error} &= (\text{Critical value}) \times (\text{Standard error}) \\ &= (1.96) \times (0.15) \\ &= 0.294\end{aligned}$$

EXAMPLE

- Step 4: Obtain the confidence interval : $\bar{X} \pm margin\ of\ error$

  - Lowerbound = Point estimate ($\bar{X}$) – margin of error
  - Upperbound = Point estimate ($\bar{X}$) + margin of error

  - Lowerbound = $6.7 - 0.294 = 6.406$
  - Upperbound = $6.7 + 0.294 = 6.994$

  - 95% confidence interval of the population mean
    = [6.406, 6.994]

INTERPRETATION

- The confidence level describes the uncertainty associated with a sampling method. Suppose we used the same sampling method to select different samples and to compute a different interval estimate for each sample. Some interval estimates would contain the true population parameter and some would not.

- A 95% confidence level indicates that we would expect 95% of the interval estimates to contain the population parameter; A 95% confidence level means that 95% of the intervals would contain the parameter.

INTERPRETATION

- In our example, the 95% confidence interval of the population mean satisfaction score of the course is [6.406, 6.994].

- We have 95% confidence that the interval [6.406, 6.994] would contain the population mean. (If a large number of repeated samples of size 100 were taken from this population and a 95% confidence interval were obtained for each sample, 95% of the confidence intervals would contain the population mean.)

EXERCISE QUESTIONS

- A researcher hypothesized that individuals suffering from Alzheimer's Disease may spend less time per night in the deeper stages of sleep. A sample of 81 Alzheimer's patients were tested to assess the amount of time in stage IV sleep. The sample produced a mean of 48 minutes. The population variance of time spent in the stage IV sleep is known to be 14 squared minutes.

- Compute a 90%confidence interval for the population average amount of time in stage IV sleep of Alzheimer's patients, and interpret it.

- Compute a 99%confidence interval for the population average amount of time in stage IV sleep of Alzheimer's patients, and interpret it.

- Compare the 90% and the 99% confidence intervals that you obtained. Which one is wider?

SUMMARY

- Estimation of a population mean
  - Point estimation : unbiased estimator
  - Interval estimation (confidence interval)
    - Margin of error is determined by two things:
      - Confidence level
      - Standard error of the point estimate (sample-to-sample variation of the point estimate)
    - Margin of error = (critical value) x (standard error)