Chapter 1

Estimation

LO 1.1: Discuss point estimators and their desirable properties.

LO 1.2: Explain an interval estimator.

LO 1.3: Calculate a confidence interval for the population mean when the population standard deviation is known.

LO 1.4: Describe the factors that influence the width of a confidence interval.

LO 1.5: Discuss features of the t distribution.

LO 1.6: Calculate a confidence interval for the population mean when the population standard deviation is not known.

LO 1.7: Calculate a confidence interval for the population proportion.

LO 1.8: Select a sample size to estimate the population mean and the population proportion.

1.1 Point Estimators and Their Properties

> LO 1.1 Discuss point estimators and their desirable properties.

1.1.1 Point Estimator

- A function of the random sample used to make inferences about the value of an unknown population parameter.

- For example, $\bar{X}$ is a point estimator for μ and $\bar{P}$ is a point estimator for p.

1.1.2   Point Estimate

- The value of the point estimator derived from a given sample.

- For example, $\bar{x} = 96.5$ is a point estimate of the mpg for all ultra-green cars.

1.1.3   Properties of Point Estimators

- Unbiased – an estimator is unbiased if its expected value equals the unknown population parameter being estimated.

- Efficient – an unbiased estimator is efficient if its standard error is lower than that of other unbiased estimators.

- Consistent – an estimator is consistent if it approaches the unknown population parameter being estimated as the sample size grows larger.

1.2   Confidence Interval of the Population Mean When σ Is Known

LO 1.2 **Explain an interval estimator.**

- Confidence Interval – provides a range of values that, with a certain level of confidence, contains the population parameter of interest.

  - Also referred to as an interval estimate.

- Construct a confidence interval as: Point estimate $\pm$ Margin of error.

  - Margin of error accounts for the variability of the estimator and the desired confidence level of the interval.

1.2.1   Constructing a Confidence Interval for μ When σ is Known

LO 1.3 **Calculate a confidence interval for the population mean when the population standard deviation is known.**

- Consider a standard normal random variable:

$$P(-1.96 \le Z \le 1.96) = 0.95$$

- Because of (**??**), we get:

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

- Which, after algebraically manipulating, is equal to:

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \tag{1.1}$$

- Note that (1.1) implies there is a 95% probability that the sample mean $\bar{X}$ will fall within the interval $\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$.

  - Thus, if samples of size n are drawn repeatedly from a given population, 95% of the computed sample means, ___, will fall within the interval and the remaining 5% will fall outside the interval.

- Since we do not know μ, we cannot determine if a particular $\bar{x}$ falls within the interval of not.

  - However, we do know that $\bar{X}$ will fall within the interval $\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$ iff μ falls within the interval $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$.

- This will happen 95% of the time given the interval construction. Thus, this is a 95% confidence interval for the population mean.

- Level of significance (i.e., probability of error) = α.

- Confidence coefficient = $1 - \alpha \Rightarrow \alpha = 1-$ confidence coefficient.

- A $100(1 - \alpha)\%$ confidence interval of the population mean μ when the standard deviation σ is known is computed as

$$\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \tag{1.2}$$

  or equivalently

$$\left[\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \tag{1.3}$$

- $z_{\frac{\alpha}{2}}$ is the z-value associated with the probability of $\frac{\alpha}{2}$ being in the upper-tail.

- Confidence Intervals:

  - 90%, $\alpha = 0.10$, $\frac{\alpha}{2} = 0.05$, $z_{0.05} = 1.645$.
  - 95%, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $z_{0.025} = 1.96$.
  - 99%, $\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$, $z_{0.005} = 2.575$.

1.2.2 Interpreting a Confidence Interval

- Interpreting a confidence interval requires care.

- Incorrect: the probability that μ falls in the interval is 0.95.

- Correct: If numerous samples of size n are drawn from a given population, then 95% of the intervals formed by the ____ $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{n}$ will contain μ.

  - Since there are many possible samples, we will be right 95% of the time, thus giving us 95% confidence.

1.2.3 The Width of a Confidence Interval

LO 1.4 Describe the factors that influence the width of a confidence interval.

- Margin of Error: $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

- Confidence Interval Width: $2\left(z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

- The width of the confidence interval is influenced by the:

  - Sample size n,
  - Standard deviation σ, and
  - Confidence level $100(1-\alpha)\%$.

1.2.4 Summary of the t_{df} Distribution

- Bell-shaped and symmetric around 0 with asymptotic tails (the tails get closer and closer to the horizontal axis, but never touch it).

- Has slightly broader tails than the z distribution.

- Consists of a family of distributions where the actual shape of each one depends on the df. As df increases, the t_{df} distribution becomes similar to the z distribution; it is identical to the z distribution when $df \to \infty$.

1.3 Confidence Interval of the Population Mean When σ Is Unknown

1.3.1 The t-Distribution

LO 1.5 Discuss features of the t distribution.

- If repeated samples of size n are taken from a normal population with a finite variance, then the statistic T follows the t-distribution __ with $n-1$ degrees of freedom, __.

- Degrees of freedom – determines the extent of the broadness of the tails of the distribution; the fewer the degrees of freedom, the broader the tails.

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \tag{1.4}$$

1.3.2　Constructing a Confidence Interval for μ When σ Is Unknown

> **LO 1.6 Calculate a confidence interval for the population mean when the population standard deviation is not known.**

- A $100(1-\alpha)\%$ confidence interval of the population mean μ when the population standard deviation σ is not known is ----

$$\bar{x} \pm t_{\alpha/2,df}\frac{s}{\sqrt{n}} \tag{1.5}$$

or equivalently

$$\left[\bar{x} - t_{\alpha/2,df}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,df}\frac{s}{\sqrt{n}}\right] \tag{1.6}$$

where s is the sample standard deviation.

1.4　Confidence Interval of the Population Proportion

> **LO 1.7 Calculate a confidence interval for the population proportion.**

- Let the parameter p represent the proportion of successes in the population, where success is defined by a particular output.

  − $\bar{p}$ is the point estimator of the population proportion p.

- By the central limit theorem, $\bar{P}$ can be approximated by a normal distribution for large samples (i.e., $np \geq 5$ and $n(1-p) \geq 5$).

- Thus, a $100(1-\alpha)\%$ confidence interval of the population proportion is

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \text{ or } \left[\bar{p} - z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] \tag{1.7}$$

where $\bar{p}$ is used to estimate the population parameter p.

1.5　Selecting a Useful Sample Size

> **LO 1.8 Select a sample size to estimate the population mean and the population proportion.**

- Precision in interval estimates is implied by a low margin of error.

- The larger n reduces the margin of error for the interval estimates.

- How large should the sample size by for a given margin of error?

1.5.1   Selecting n to Estimate μ

- Consider a confidence interval for μ with a known σ and let D denote the desired margin or error.

- Since

$$D = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \tag{1.8}$$

  we may rearrange to get

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{D} \right)^2. \tag{1.9}$$

- If σ is unknown, estimate it with $\hat{\sigma}$.

- For a desired margin of error D, the minimum sample size n required to estimate a $100(1 - \alpha)\%$ confidence interval of the population mean μ is

$$n = \left(\frac{z_{\frac{\alpha}{2}} \hat{\sigma}}{D} \right)^2. \tag{1.10}$$

  where $\hat{\sigma}$ is a reasonable estimate of σ in the planning stage.

1.5.2   Selecting n to Estimate p

- Consider a confidence interval for p and let D denote the desired margin of error.

- Since

$$D = z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \tag{1.11}$$

  (where $\bar{p}$ is the sample proportion), we may rearrange to get

$$n = \left(\frac{z_{\alpha/2}}{D} \right)^2 \bar{p}(1 - \bar{p}) \tag{1.12}$$

- Since $\bar{p}$ comes from a sample, we must use a reasonable estimate of p, that is, $\hat{p}$.

- For a desired margin of error D, the minimum sample size n required to estimate a $100(1 - \alpha)\%$ confidence interval of the population proportion p is

$$n = \left(\frac{z_{\alpha/2}}{D} \right)^2 p(1 - p) \tag{1.13}$$

  where $\hat{p}$ is a reasonable estimate of p in the planning stage.