Chapter 7. Notes

Estimating Parameters and Determining Sample Sizes

- **Sample Size:** We should know how to find the sample size necessary to estimate a population proportion.
- Confidence Interval
  - A **confidence interval** (or **interval estimate**) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.
  - The **confidence level** is the probability $1 - \alpha$ (such as 0.95, or 95%) that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. (The confidence level is also called the **degree of confidence**, or the **confidence coefficient**.)

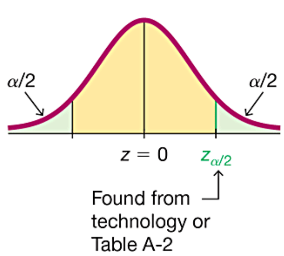
The following table shows the relationship between the confidence level and the corresponding value of α. The confidence level of 95% is the value used most often.

Most Common Confidence Levels	**Corresponding Values of α**
90% (or 0.90) confidence level:	$\alpha = 0.10$
95% (or 0.95) confidence level:	$\alpha = 0.05$
99% (or 0.99) confidence level:	$\alpha = 0.01$

Correct interpretations of the confidence interval $0.405 < p < 0.455$:

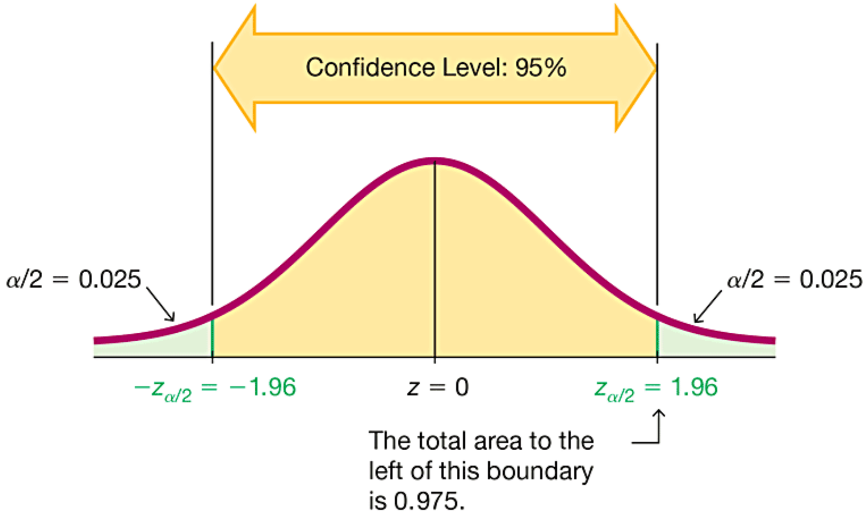
"We are 95% confident that the interval from 0.405 to 0.455 actually does contain the true value of the population proportion *p*."

A **critical value** is the number on the borderline separating sample statistics that are significantly high or low from those that are not significant.



Note that when finding the critical *z* score for a 95% confidence level, we use a cumulative left area of 0.9750 (**not** 0.95). Think of it this way:

This is our confidence level:		The area in *both* tails is:		The area in the *right* tail is:		The cumulative area from the left, excluding the right tail, is:
95%	→	$\alpha = 0.05$	→	$\alpha/2 = 0.025$	→	$1 - 0.025 = 0.975$



This is the most common critical value, and it is listed with two other common values in the table that follows.

Confidence level	α	Critical Value, $z_{\frac{\alpha}{2}}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

Confidence Interval for Estimating a Population Proportion *p*: Confidence Interval Estimate of *p:*

$$\hat{p} - E < p < \hat{p} + E \quad \text{where} \quad E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

The confidence interval is often expressed in the following formats:

$$\hat{p} \pm E \quad \text{or} \quad (\hat{p} - E,\ \hat{p} + E)$$

Determining Sample Size: Finding the Sample Size Required to Estimate a Population Proportion: Requirements:

When an estimate $\hat{p}$ in known:

$$n = \frac{\left[z_{\frac{\alpha}{2}}\right]^2 \hat{p}\hat{q}}{E^2}$$

When no estimate $\hat{p}$ is known:

$$n = \frac{\left[z_{\frac{\alpha}{2}}\right]^2 0.25}{E^2}$$

Confidence Interval for Estimating a Population Mean with σ Not Known: Requirements

1. The sample is a simple random sample.
2. Either or both of these conditions are satisfied: The population is normally distributed or $n > 30$.

Confidence Interval for Estimating a Population Mean with σ Not Known: Confidence Interval:

Margin of Error:

$$E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \quad \text{(\textbf{Use df} } = n - 1\text{)}$$

Procedure for Constructing a Confidence Interval for μ:

Using the value of the calculated margin of error E and the sample mean $\bar{x}$, substitute the values in one of the formats for CI:

$$\bar{x} - E < \mu < \bar{x} + E \text{ or } \bar{x} \pm E \text{ or } (\bar{x} - E, \bar{x} + E).$$

With an **original set of data** values, round the confidence interval limits to one more decimal place than used for the original set of data, but when using the **summary statistics** of n, $\bar{x}$, and s, round the confidence interval limits to the same number of decimal places used for the sample mean.

Finding the Sample Size Required to Estimate a Population Mean: Sample Size

The required sample size is found by

$$n = \left[\frac{z_{\frac{\alpha}{2}} \sigma}{E} \right]^2$$

If the computed sample size n is not a whole number, round the value of n up to the next **larger** whole number.

Estimating a Population Mean When σ Is Known

If we somehow do know the value of s, the confidence interval is constructed using the standard normal distribution instead of the Student t distribution, so the same procedure can be used with this margin of error:

Margin of Error: $E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ (used with known σ).

The confidence interval is given by:

$$\bar{x} - E < \mu < \bar{x} + E \text{ or } \bar{x} \pm E \text{ or } (\bar{x} - E, \bar{x} + E).$$

Choosing between Student *t* and *z* (Normal) Distributions:

Conditions	**Method**
σ not known and normally distributed population **or** σ not known and $n > 30$	Use student *t* distribution
σ known and normally distributed population **or** σ known and $n > 30$ (In reality, σ is rarely known.)	Use normal (*z*) distribution.
Population is not normally distributed and $n \leq 30$.	Use the bootstrapping method por a nonparametric method.