CHAPTER 7 Estimates, Sample Sizes, and Confidence Intervals

The following notes were taken from the PowerPoint slides for this chapter and supplemented as appropriate to make the notes more complete for this course.

GOALS
When you have completed this chapter, you will be able to:
1. Define what is meant by a point estimate.
2. Define the term level of confidence.
3. Construct a confidence interval for the population proportion.
4. Construct a confidence interval for the population mean when the population standard deviation is known.
5. Construct a confidence interval for the population mean when the population standard deviation is unknown.
6. Determine the sample size for attribute and variable sampling.

Point Estimates and Interval Estimates
A point estimate is a single value (statistic) used to estimate a population value (parameter).
A confidence interval is a range of values within which the population parameter is expected to occur. The factors that determine the width of a confidence interval are:
1. The sample size, \( n \).
2. The proportion or variability in the population.
3. The desired level of confidence.

Choosing the Proper Test Statistic – z-test vs. t-test:
1. For confidence intervals of the proportion, we use the z-test. 
   \[
   \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
   \]
2. For confidence intervals of the mean: We use the z-test, 
   \[
   \bar{X} \pm z \frac{\sigma}{\sqrt{n}}
   \]
   a) If the population standard deviation is known and 
   b) Either the population is a normal distribution or the sample is greater than 30.
3. For confidence intervals of the mean: If the conditions of 2 (a) and 2 (b) are not met, we use the t-test, 
   \[
   t = \frac{\bar{x} - \mu}{s/\sqrt{n}}
   \]

Interval Estimates
a. An interval estimate states the range within which a population parameter probably lies.
b. The interval within which a population parameter is expected to occur is called a confidence interval.
c. The two confidence intervals that are used extensively are the 95% and the 99%.
d. For a 95% confidence interval, about 95% of the similarly constructed intervals will contain the parameter being estimated. Also 95% of the sample means for a
specified sample size will lie within 1.96 standard deviations of the hypothesized population mean.

e. For the 99% confidence interval, 99% of the sample means for a specified sample size will lie within 2.575 standard deviations of the hypothesized population mean.

**Standard Error of the Sample Means**

The standard error of the sample mean is the standard deviation of the sampling distribution of the sample means.

When **σ is known**, the standard error of the sample is computed by \( \sigma_x = \frac{\sigma}{\sqrt{n}} \)

- \( \sigma_x \) is the symbol for the standard error of the sample mean.
- \( \sigma \) is the standard deviation of the population.
- \( n \) is the size of the sample.

When **σ is not known** and \( n > 30 \), the standard deviation of the sample, designated \( s \), is used to approximate the population standard deviation. The formula for the standard error of the sample is:

\[ s_x = \frac{s}{\sqrt{n}} \]

**95% and 99% Confidence Intervals for \( \mu \)**

The 95% and 99% confidence intervals for are constructed as follows when:

**95% CI** for the population mean is given by \( \overline{X} \pm 1.96 \frac{s}{\sqrt{n}} \)

**99% CI** for the population mean is given by \( \overline{X} \pm 2.575 \frac{s}{\sqrt{n}} \)

**Constructing General Confidence Intervals for \( \mu \)**

In general, a confidence interval for the mean is computed by: \( \overline{X} \pm z \frac{s}{\sqrt{n}} \)

**EXAMPLE 3**

The Dean of the Business School wants to estimate the mean number of hours worked per week by students. A sample of 49 students showed a mean of 24 hours with a standard deviation of 4 hours. What is the population mean?

The value of the *population mean* is not known. Our best estimate of this value is the sample mean of 24.0 hours. This value is called a point estimate.

Find the 95 percent confidence interval for the population mean.

\[ \overline{X} \pm 1.96 \frac{s}{\sqrt{n}} = 24.00 \pm 1.96 \frac{4}{\sqrt{49}} = 24.00 \pm 1.12 \]
The confidence limits range from 22.88 to 25.12. About 95 percent of the similarly constructed intervals include the population parameter.

Confidence Interval for a Population Proportion

The confidence interval for a population proportion is estimated by: \( \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \)

EXAMPLE 4
A sample of 500 executives who own their own home revealed 175 planned to sell their homes and retire to Arizona. Develop a 98% confidence interval for the proportion of executives that plan to sell and move to Arizona.

\[
0.35 \pm 2.33 \sqrt{\frac{(0.35)(0.65)}{500}} = 0.35 \pm 0.0497
\]

Finite-Population Correction Factor
A population that has a fixed upper bound is said to be finite.
For a finite population, where the total number of objects is \( N \) and the size of the sample is \( n \), the following adjustment is made to the standard errors of the sample means and the proportion: \( \sqrt{\frac{N-n}{N-1}} \)

Standard error of the sample means: \( \sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \)

Finite-Population Correction Factor

Standard error of the sample proportions: \( \sigma_p = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \)

This adjustment is called the finite-population correction factor. If \( n/N < .05 \), the finite-population correction factor is ignored.

EXAMPLE 5
Given the information in EXAMPLE 3, construct a 95% confidence interval for the mean number of hours worked per week by the students if there are only 500 students on campus. [mean = 24, \( \sigma = 4 \), \( n = 49 \), and \( N = 500 \)]

Because \( n/N = 49/500 = .098 \) which is greater than .05, we use the finite population correction factor.

\[
\overline{X} \pm \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 24 \pm \frac{4}{\sqrt{49}} \sqrt{\frac{500-49}{500-1}} = 24.00 \pm 1.0648
\]
Selecting a Sample Size

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population. They are:

1. The degree of confidence selected.
2. The maximum allowable error.
3. The variation in the population.

Sample Size for Means

To find the sample size for a variable use

$$n = \left( \frac{z \cdot s}{E} \right)^2$$

where $E$ is the allowable error, $z$ is the $z$-value corresponding to the selected level of confidence, and $s$ is the sample deviation of the pilot survey.

EXAMPLE 6

A consumer group would like to estimate the mean monthly electricity charge for a single family house in July within $5 using a 99 percent level of confidence. Based on similar studies the standard deviation is estimated to be $20.00. How large a sample is required?

$$n = \left( \frac{(2.58)(20)}{5} \right)^2 = 106.5024$$

We round up to the next whole number. So use $n = 107$.

Sample Size for Proportions

The formula for determining the sample size in the case of a proportion is:

$$n = \frac{p(1-p)}{E^2} \left( \frac{z}{E} \right)^2$$

where $p$ is the estimated proportion, based on past experience or a pilot survey; $z$ is the $z$ value associated with the degree of confidence selected; $E$ is the maximum allowable error the researcher will tolerate.

EXAMPLE 7

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = .30(1-.30) \left( \frac{1.96}{.03} \right)^2 = 896.37$$

We round up to the next whole number. So use $n = 897$. 