Chapter 8: Confidence Intervals

Introduction

There are two major techniques for classical statistical inference: confidence intervals and hypothesis testing.

A **point estimate** is a single value used to estimate a population parameter.

- The sample proportion $\hat{p}$ is the best *point estimate* of the population proportion, p.
- The sample mean $\bar{x}$ is the best *point estimate* of the population mean, μ.

We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct interval estimates, called confidence intervals.

A **confidence interval** is an interval of values computed from sample data that is likely to include the unknown value of a population parameter. There is no guarantee that a given confidence interval does capture the parameter, but there is a predictable probability of success.

A confidence interval is always accompanied by a **confidence level**, which tells us that, after repeated sampling, the confidence interval contains the parameter of interest.

- Expressed as $1 - \alpha$, often as a percentage.
- α is referred to as our significance level.

Theory: if we select many different samples of size n from the same population and constructed the corresponding confidence intervals, 95% of them would contain the parameter.

The **margin of error** depends on the confidence level or percentage of confidence and the standard error of the mean.

$$\text{point estimate} \ \pm \ \text{margin of error}$$

Interpretation: "We are [confidence level] confident the population proportion is between [lower bound] and [upper bound]."

Notes:

- The half-width of a confidence interval is often called the **margin of error,** E.

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$E = t^* \left(\frac{s}{\sqrt{n}}\right)$$

- As the confidence level increases, the margin-of-error and width of the confidence interval also increase. The midpoint of the interval remains at the sample proportion, ($\hat{p}$), or sample mean, ($\bar{x}$), regardless of the confidence level.

- A larger sample size produces a narrower confidence interval whenever other factors remain the same. In this case, we generally say our estimate of the population parameter is more *precise*.

- Confidence intervals estimate the value of a population *parameter*; they do not estimate the value of a sample statistic or of an individual observation.

8.3: Population Proportion

To construct a confidence interval for a single unknown population proportion, p, we need a point estimate for p and the margin of error, E.

- The sample proportion, $\hat{p}$, is the best *point estimate* of the population proportion, p.
- The *margin of error*, E, is the critical value times the standard deviation for the sample proportion, $z^*\sqrt{\dfrac{p(1-p)}{n}}$.
  - z^* represents the critical value from the standard normal distribution for the confidence level desired.

Confidence interval for a population proportion, p:

$$\text{point estimate} \ \pm \ \text{margin of error}$$

To find confidence intervals with varying confidence levels, we will change the **critical value**, which is denoted by z^*, and can be found using the standard normal distribution (z-table).

ConfidenceLevel	Critcal Value, z^*	Confidence Interval
90%	1.645	$\hat{p} \pm 1.645$ standard error
95%	1.96	$\hat{p} \pm 1.96$ standard error
98%	2.33	$\hat{p} \pm 2.33$ standard error
99%	2.575	$\hat{p} \pm 2.575$ standard error

NOTE: Your textbook uses $z_{\alpha/2}$ as the critical value, where we use z^*.

Interpretation: "We are [confidence level] confident the population mean is between [lower bound] and [upper bound]."

Conditions:

1. Representative sample
2. X, the number of successes, follows a binomial distribution
3. Both $n\hat{p}$ and $n\hat{q}$ are at least 5 (at least 5 successes and 5 failurs)

Working Backwards to Find the Margin of Error or Sample Mean

When we calculate a confidence interval, we find the sample mean, calculate the margin of error, and use them to calculate the confidence interval. However, sometimes when we read statistical studies, the study may state the confidence interval only. If we know the confidence interval, we can work backwards to find both the error bound and the sample mean.

Finding the Margin of Error

- From the upper value for the interval, subtract the sample mean,
- OR, from the upper value for the interval, subtract the lower value. Then divide the difference by two.

Finding the Sample Mean

- Subtract the error bound from the upper value of the confidence interval,
- OR, average the upper and lower endpoints of the confidence interval.

Notice that there are two methods to perform each calculation. You can choose the method that is easier to use with the information you know

Example 1. Suppose you have the following confidence interval: $(10.21, 10.59)$.

(a) Find the margin of error.

(b) Find the point estimate.

To find z^*, note that the standard normal curve can be partitioned into three sections:

1. The area between $-z^*$ and z^*, which is CL, the confidence level.
2. The area below $-z^*$, which is half of the area not betweeen $-z^*$ and z^*. This area is $(1 - CL)/2 = \alpha/2$.
3. The area above $+z^*$, which by symmetry equals the area below $-z^*$. This area is $(1 - CL)/2 = \alpha/2$.

One strategy for determining z^*, or $z_{\alpha/2}$, for a general confidence level is to use the fact that $-z^*$ is the z-score that has area $(1 - C)/2$ below it. This can easily be read from the z-table provided. Search within the body of the table and identify the corresponding value of $-z^*$.

Example 2. Find the critical value for a confidence interval in which we are 88% confident.

Example 3. Find the critical value for a confidence interval in which we are 80% confident.

Example 4. In April 1998, the Marist Institute for Public Opinion surveyed 883 randomly selected American adults about allergies. According to a report posted at the Institute's website, 36% of the sample answered "yes" to the question "Are you allergic to anything?"

(a) Construct a 95% confidence interval to estimate the population parameter p = proportion of all American adults who are allergic to something.

- point estimate $\hat{p} = 0.36$
- sample size $n = 883$
- critical value is $z^* = 1.96$ (to achieve 95% confidence)
- margin of error $E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.36(1-0.36)}{883}} = 0.03166$

The 95% confidence interval is

$$\hat{p} \pm E$$

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.36 \pm 0.03166$$

$$(0.3283, 0.3917)$$

(b) Interpret the interval.

We are **95% confident** the **population proportion** of American adults who are allergic to something is between **32.83% and 39.17%**.

(c) Suppose a researcher claimed that less than 40% of the population would answer "yes." Address this claim based on our interval.

Because the value 0.40 does NOT lie in the interval, and our interval is LESS THAN 0.40, we can conclude the claim is true.

(d) Address all assumptions and the valididty of your results.

- sampling methodology is sound
- X follows a binomial distribution with
  $n\hat{p} = 883 * 0.36 = 317.88 \geq 5$ and
  $n\hat{q} = 883 * (1-0.36) = 565.12 \geq 5$.
- The needed assumption that $\hat{p} \dot{\sim}$ normal is confirmed.

Example 5. A six-sided die is rolled 50 times, and 4 of them produce a 1.

(a) Construct a 90% confidence interval on the proportion of times the die produces a 1.

(b) Interpret the interval.

(c) Address the claim that the die is unbalanced.

(d) Address all assumptions and the validity of your results.

Example 6. Suppose you have the following confidence interval: $(0.34, 0.42)$.

(a) Find the margin of error.

(b) Find the point estimate.

8.2: A Single Population Mean Using the Student t Distribution

To construct a confidence interval for a single unknown population mean, μ, **where the population standard deviation is UNKNOWN**, we need a point estimate for μ and the margin of error, E.

We use the t-distribution for means (instead of the z-distribution) if:

- if we do NOT know the population standard deviation, σ

- The sample mean, $\bar{x}$, is the best *point estimate* of the population mean, μ.
- The *margin of error*, E, is the critical value times the standard deviation for the sample mean, $t^* \left(\frac{\sigma}{\sqrt{n}}\right)$.
  - t^* represents the critical value from the Student's t distribution for the confidence level desired, with $n-1$ degrees of freedom.
- The *degrees of freedom* is the number of sample observations that can vary after certain restrictions have been imposed. For the t distribution, the degrees of freedom is equal to $n-1$.

To use the t-table, compute the degrees of freedom (df), choose a confidence level (look at two-tailed row), and then look in the row and column for these values in the table.

NOTE: Your textbook uses $t_{\alpha/2}$ as the critical value, where we use t^*.

Confidence interval for a population mean, μ (t-interval):

$$\text{point estimate} \ \pm \ \text{margin of error}$$

Conditions for Normality

1. Representative Sample
2. One of the following:
   (a) The population must be normally distributed, $X \sim$ normal, OR
   (b) The sample size needs to be *large* enough, $n \geq 30$, OR
   (c) Check the Q-Q plot and boxplot.

Interpretation: "We are [confidence level] confident the population mean is between [lower bound] and [upper bound]."

Properties of the Student's t Distribution

- There are an infinite number of t distributions, one for each possible value for degrees of freedom.
- The t distribution has the same symmetric bell shape as the z distribution, but it reflects the greater variability that is expected with estimating σ.
- The t distribution has a mean of 0.
- The standard deviation of the t distribution varies with the sample size, but is always greater than 1 (unlike the z).
- As the sample size n gets larger, the t distribution gets closer to the z distribution.

Example 7. Find the t^* critical value for the following:

1. 95% confidence, $n = 15$.

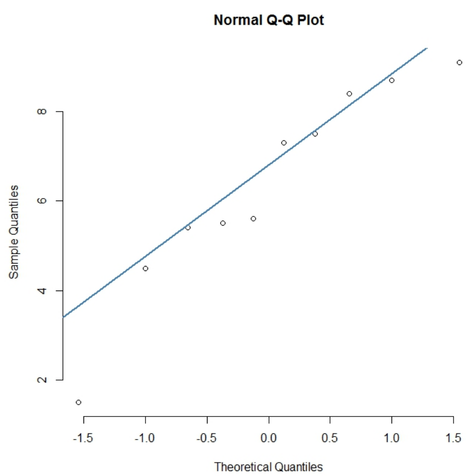
2. 99% confidence, $n = 43$

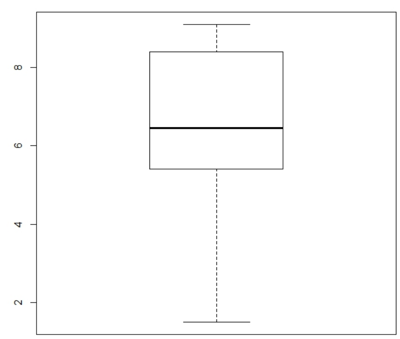
3. 80% confidence, $n = 30$

Example 8. Listed below are amounts of arsenic (μg, or micrograms, per serving) in samples of brown rice from California (based on data from the Food and Drug Administration). Treat the sample as a simple random sample for arsenic in all brown rice from California. Construct a 95% confidence interval for the mean arsenic level of brown rice in California. Interpret this interval.

Step 1 Because the sample is small, we verify that the sample data come from a population that is normally distributed with no outliers.

5.4
5.6
8.4
7.3
4.5
7.5
1.5
5.5
9.1
8.7





Step 2

Step 3

Step 4

Step 5

Mixed Problems

Example 9. A physician wants to test a medicine to lower blood glucose concentrations in diabetic patients. The drug is considered "successful" if it can lower blood glucose concentrations below 110mg/dL. He uses the drug on 30 patients, and at the end of the trial the sample mean is 102 mg/dL with a sample standard deviation of 3.2. Assume values are normally distributed.

1. Construct a 90% confidence interval for the true value of the population mean blood glucose concentration.
   (a) Check the conditions for the validity of the test.
   (b) Determine the critical value.
   (c) Determine the point estimate and margin of error.
   (d) Construct the confidence interval.
   (e) Interpret the interval found in part (d).
2. The physician claims the drug is successful. Is this claim reasonable?

Example 10. The Dallas Stars active roster has 24 players on it, and the average player weight (in pounds) is 200.2, while the standard deviation (in pounds) is 18.2.

(a) Assuming that the data is somewhat symmetric and treating the Stars as a representative sample of NHL players, construct a 95% confidence interval on the average weight of hockey players.

(b) Interpret the interval.

(c) Address the claim that the mean weight of all hockey players is over 195 pounds.

(d) Address all assumptions and the validity of your results.

Example 11. A Harris Interactive poll conducted during January 2008 found that 944 of 1748 adult Americans 18 years or older who do not have a tattoo believe that individuals with tattoos are more rebellious.

(a) Obtain a point estimate for the proportion of adult Americans without tattoos who believe individuals with tattoos are more rebellious.

(b) Verify that the requirements for constructing a confidence interval for p are satisfied.

(c) Construct a 90% confidence interval for the proportion of adult Americans without tattoos who believe individuals with tattoos are more rebellious.

(d) Construct a 99% confidence interval for the proportion of adult Americans without tattoos who believe individuals with tattoos are more rebellious.

(e) What is the effect of increasing the level of confidence on the width of the interval?

Example 12. Suppose you have the following confidence interval: $(0.247, 0.304)$.

(a) Find the margin of error.

(b) Find the point estimate.

Example 13. Suppose you have the following confidence interval: $(16.83, 18.27)$.

(a) Find the margin of error.

(b) Find the point estimate.

