Chapter 9: Testing a Claim

Objectives: Students will:

Explain the logic of significance testing.
List and explain the differences between a null hypothesis and an alternative hypothesis.
Discuss the meaning of statistical significance.
Use the Inference Toolbox to conduct a large sample test for a population mean.
Compare two-sided significance tests and confidence intervals when doing inference.
Differentiate between statistical and practical "significance."
Explain, and distinguish between, two types of errors in hypothesis testing.
Define and discuss the power of a test.
Conduct *one-sample* and *paired data t significance tests*.
Explain the differences between the one-sample confidence interval for a population proportion and the one-sample significance test for a population proportion.
Conduct a *significance test for a population proportion*.

AP Outline Fit:
IV. Statistical Inference: Estimating population parameters and testing hypotheses (30%–40%)

B. Tests of significance

1. Logic of significance testing, null and alternative hypotheses; *P*-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power
2. Large sample test for a proportion
4. Test for a mean

What you will learn:

A. Significance Tests for μ (σ known)
  1. State the null and alternative hypotheses in a testing situation when the parameter in question is a population mean μ.
  2. Explain in nontechnical language the meaning of the P-value when you are given the numerical value of P for a test.
  3. Calculate the one-sample z-statistic and the P-value for both one-sided and two-sided tests about the mean μ of a Normal population
  4. Explain Type I error, Type II error, and power in a significance-testing problem.
  5. Assess statistical significance at standard levels α by comparing *P* to α.
  6. Recognize that significance testing does not measure the size or importance of an effect.
  7. Recognize when you can use the *z* test and when the data collection design or a small sample from a skewed population makes it inappropriate.

B. One-Sample *t* Test for μ (σ unknown)
  1. Carry out a *t* test for the hypothesis that a population mean μ has a specified value against either a one-sided or a two-sided alternative. Use Table C of *t* critical values to approximate the *P*-value or carry out a fixed α test.
  2. Recognize when the *t* procedures are appropriate in practice, in particular that they are quite robust against lack of Normality but are influenced by skewness and outliers.
  3. Also recognize when the design of the study, outliers, or a small sample from a skewed distribution make the *t* procedures risky.
  4. Recognize paired data and use the *t* procedures to perform significance tests for such data.

C. Inference about One Proportion
  1. Use the *z* statistic to carry out a test of significance for the hypothesis $H_0: p = p_0$ about a population proportion *p* against either a one-sided or a two-sided alternative.
  2. Check that you can safely use the one-proportion *z* test in a particular setting.

Chapter 9: Testing a Claim

Section 9.1: Significance Tests: The Basics

Objectives: Students will:

State appropriate hypothesis for a significance test about a population parameter
Interpret a P-value in context
Make an appropriate conclusion for a significance test
Interpret a Type I error and a Type II error in context. Give a consequence of each error in a given setting

Vocabulary:

Null Hypothesis – H_0, the claim that we weigh evidence against
Alternative Hypothesis – H_a, or H_1, the claim that we are trying to find evidence for
One-sided alternative hypothesis – states that a parameter is greater than or less than the null value
Two-sided alternative hypothesis – states that a parameter is not equal to (or different from) the null value
P-value – the probability of getting evidence for the alternative hypothesis H_a as strong as or stronger than the observed evidence when the null hypothesis H_0 is true
Significance Level – the value that we use as a boundary for deciding whether an observed result is unlikely to happen by chance alone when the null hypothesis is true
Type I error – occurs if a test rejects H_0 when H_0 is true. That is, the test finds convincing evidence that H_a is true, when it really isn't
Type II error – occurs if a test fails to rejects H_0 when H_a is true. That is, the test does not find convincing evidence that H_a is true, when it really is

Hypothesis – a statement or claim regarding a characteristic of one or more populations
Hypothesis Testing – procedure, base on sample evidence and probability, used to test hypotheses
Level of Significance – probability of making a Type I error, α
Power of the test – value of 1 – β
Power curve – a graph that shows the power of the test against values of the population mean that make the null hypothesis false.

Key Concepts:

Hypotheses: Null H_0 & Alternative H_a

Null and Alternative Hypotheses

The statement being tested in a significance test is called the **null hypothesis.** The significance test is designed to assess the strength of the evidence *against* the null hypothesis. Usually the null hypothesis is a statement of "no effect," "no difference," or no change from historical values.

The claim about the population that we are trying to find evidence *for* is the **alternative hypothesis.**

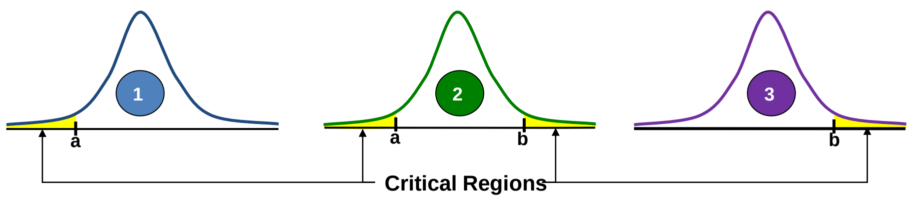
- Think of the null hypothesis as the status quo
- Think of the alternative hypothesis as something has changed or is different than expected
- We cannot prove the null hypothesis! We only can find enough evidence to reject the null hypothesis or not.
- Our hypotheses will only involve population parameters (we know the sample statistics!)
- The alternative hypothesis can be
  - one-sided: $\mu > 0$ or $\mu < 0$ (which allows a statistician to detect movement in a specific direction)
  - two-sided: $\mu \neq 0$ (things have changed)
  - Read the problem statement carefully to decide which is appropriate
- The null hypothesis is usually "=", but if the alternative is one-sided, the null could be too

Chapter 9: Testing a Claim

Conditions for Significance Tests

- **SRS**
  - check problem statement; can it be assumed if not specifically stated?
- **Independence**
  - sample less than 10% of population
- **Normality**
  - Means
    - population distribution is stated as Normal
    - large enough sample size ($n \geq 30$) for CLT to apply
  - Proportions
    - $np \geq 10$ and $n(1-p) \geq 10$

Three Ways – H_o versus H_a



1. Equal versus less than (left-tailed test)
   H_0: the parameter = some value (or more)
   H_1: the parameter < some value

2. Equal hypothesis versus not equal hypothesis (two-tailed test)
   H_0: the parameter = some value
   H_1: the parameter ≠ some value

3. Equal versus greater than (right-tailed test)
   H_0: the parameter = some value (or less)
   H_1: the parameter > some value

Test Statistics

Principles that apply to most tests:

- **The test is based on a statistic that compares the value of the parameter as stated in H_0 with an estimate of the parameter from the sample data**
- **Values of the estimate far from the parameter value in the direction specified by H_a give evidence against H_0**
- **To assess how far the estimate is from the parameter, standardize the estimate. In many common situations, the test statistic has the form:**

$$\text{test statistic} = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard deviation of the estimate (ie SE)}}$$

Chapter 9: Testing a Claim

Hypothesis Testing Approaches

- **P-Value**
  - Logic: Assuming H_0 is true, if the probability of getting a sample mean *as extreme or more extreme* than the one obtained is small, then we reject the null hypothesis (accept the alternative).
    - Small P-values are evidence against H_0 (observed value is unlikely to occur if H_0 is true)
    - Large P-values fail to give evidence against H_0
- **Confidence Intervals**
  - Logic: If the sample mean lies in the confidence interval about the status quo, then we fail to reject the null hypothesis
- **Classical (Statistical Significance)**
  - Logic: If the sample mean is too many standard deviations from the mean stated in the null hypothesis, then we reject the null hypothesis (accept the alternative)

Hypothesis Testing – Four Outcomes

		Population Truth	
		H_0 is True	H_a is True
Decision based on sample	Reject H_0	Type I Error	Correct Conclusion
	Do Not Reject H_0	Correct Conclusion	Type II Error

H_0: the defendant is innocent
H_1: the defendant is guilty

decrease α → increase β
increase α → decrease β

Type I Error: convict an innocent person **α = P(T1E)**
Type II Error: let a guilty person go free **β = P(T2E)**

Note: a defendant is never declared innocent; just not guilty

- We reject the null hypothesis when the alternative hypothesis is true (Correct Decision)
- We do not reject the null hypothesis when the null hypothesis is true (Correct Decision)
- We reject the null hypothesis when the null hypothesis is true (Incorrect Decision – Type I error)
- We do not reject the null hypothesis when the alternative hypothesis is true (Incorrect Decision – Type II error)
- Significance level is the same thing as the probability of a Type I error
- Based on the diagram from the previous slide we can see the inverse relationship between α and β

Power and Type II Error

Power and Type II Error

The probability that a fixed level α significance test will reject H_0 when a particular alternative value of the parameter is true is called the **power** of the test against that alternative.

The power of a test against any alternative is 1 minus the probability of a Type II error for that alternative; that is, power $= 1 - \beta$.

- **Probability of a Type II error is β**
- **Power of the test is 1 – β**
- **P-value describes what would happen supposing the null hypothesis is true**
- **Power describes what would happen supposing that a particular alternative is true**

Chapter 9: Testing a Claim

Increasing the Power of a Test

- Increase α.
  - Testing at a higher significance level increase the chance of rejecting the alternative
- Consider a particular alternative that is farther away from μ_0
  - Values of μ (in H_a) closer to u_0 are harder to detect
- Increase the sample size.
  - More data provides more information about x-bar – better chance of distinguishing values of μ
- Decrease σ.
  - Improving the measurement process
  - Restricting attention to a subpopulation

Example 1: A manufacturer claims that there are at least two scoops of cranberries in each box of cereal

Parameter to be tested:

Test Type:

H_0:

H_a:

Example 2: A manufacturer claims that there are exactly 500 mg of a medication in each tablet

Parameter to be tested:

Test Type:

H_0:

H_a:

Example 3: A pollster claims that there are at most 56% of all Americans are in favor of an issue

Parameter to be tested:

Test Type:

H_0:

H_a:

Example: Job Satisfaction

For the job satisfaction study, the hypotheses are

H_0: $\mu = 0$

H_a: $\mu \neq 0$

Data from the 18 workers gave x-bar = 17 and sx=60. That is, these workers rated the self-paced environment, on average, 17 points higher. Researchers performed a significance test using the sample data and found a P-value of 0.2302.

a) Explain what it means for the null hypothesis to be true in this setting.

b) Interpret the *P*-value in context.

Chapter 9: Testing a Claim

Example 4: Several cities have begun to monitor paramedic response times. In one such city, the mean response time to all accidents involving life-threatening injuries last year was μ=6.7 minutes with σ=2 minutes. The city manager shares this info with the emergency personnel and encourages them to "do better" next year. At the end of the following year, the city manager selects a SRS of 400 calls involving life-threatening injuries and examines response times. For this sample the mean response time was x-bar = 6.48 minutes. Do these data provide good evidence that the response times have decreased since last year?
List hypotheses and conditions check

H_0:

H_a:

Conditions Check:
1)

2)

3)

What is the P-value associated with the data in example 4?

What if the sample mean was 6.61?

Example 5: For each α and observed significance level (p-value) pair, indicate whether the null hypothesis would be rejected.

a) α = . 05, p = .10

b) α = .10, p = .05

c) α = .01 , p = .001

d) α = .025 , p = .05

e) α = .10, p = .45

Chapter 9: Testing a Claim

Example 6: You have created a new manufacturing method for producing widgets, which you claim will reduce the time necessary for assembling the parts. Currently it takes 75 seconds to produce a widget. The retooling of the plant for this change is very expensive and will involve a lot of downtime.

H_o :

H_a:

TYPE I:

TYPE II:

Example 7: A potato chip producer wants to test the hypothesis

H_0:

H_a:

Let's examine the two types of errors that the producer could make and the consequences of each

Type I Error:

Type II Error:

Homework: Problems 1, 9, 15, 23

Section 9.2: Tests about a Population Proportion

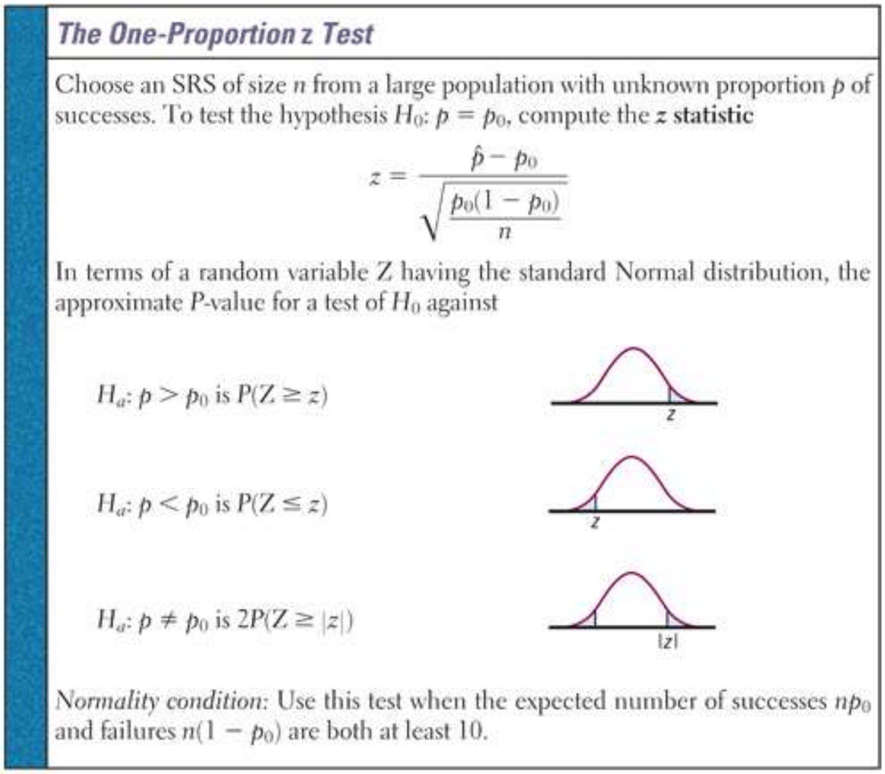
Objectives: Students will:

- State and check the Random, 10% and Large Counts conditions for performing a significance test about a population proportion.
- Calculate the standardized test statistic and P-value for a test about a population proportion
- Perform a significance test about a population proportion

Vocabulary:

Standardized test statistic – measures how far a sample statistic is from what we would expect if the null hypothesis H0 were true, in standard deviation units.

Key Concepts:



The One-Proportion z Test

Choose an SRS of size n from a large population with unknown proportion p of successes. To test the hypothesis H_0: $p = p_0$, compute the **z statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

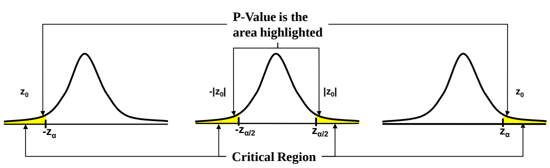
In terms of a random variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against

H_a: $p > p_0$ is $P(Z \geq z)$

H_a: $p < p_0$ is $P(Z \leq z)$

H_a: $p \neq p_0$ is $2P(Z \geq |z|)$

Normality condition: Use this test when the expected number of successes np_0 and failures $n(1 - p_0)$ are both at least 10.



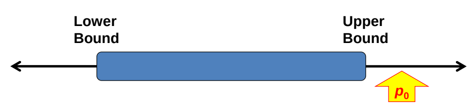
Test Statistic: $z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Reject null hypothesis, if		
P-value < α		
Left-Tailed	Two-Tailed	Right-Tailed
$z_0 < -z_\alpha$	$z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$	$z_0 > z_\alpha$

Confidence Interval Approach

Confidence Interval:

$\hat{p} - z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$ $\quad$ $\hat{p} + z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$



Reject null hypothesis, if
p_0 is not in the confidence interval

P-value associated with lower bound must be doubled!

Chapter 9: Testing a Claim

Example 1: According to OSHA, job stress poses a major threat to the health of workers. A national survey of restaurant employees found that 75% said that work stress had a negative impact on their personal lives. A random sample of 100 employees form a large restaurant chain finds 68 answered "Yes" to the work stress question. Does this offer evidence that this company's employees are different from the national average?

- Step 1: Hypothesis
  H_0:

  H_a:
- Step 2: Conditions
  SRS:

  Independence:

  Normality:
- Step 3: Calculations

  Test Statistic:
- Step 4: Interpretation

Example 2: Nexium is a drug that can be used to reduce the acid produced by the body and heal damage to the esophagus due to acid reflux. Suppose the manufacturer of Nexium claims that more than 94% of patients taking Nexium are healed within 8 weeks. In clinical trials, 213 of 224 patients suffering from acid reflux disease were healed after 8 weeks. Test the manufacturers claim at the $\alpha=0.01$ level of significance.

- Step 1: Hypothesis
  H_0:

  H_a:
- Step 2: Conditions
  SRS:

  Independence:

  Normality:
- Step 3: Calculations

  Test Statistic:
- Step 4: Interpretation

Chapter 9: Testing a Claim

Example 3: According to USDA, 48.9% of males between 20 and 39 years of age consume the minimum daily requirement of calcium. After an aggressive "Got Milk" campaign, the USDA conducts a survey of 35 randomly selected males between 20 and 39 and found that 21 of them consume the min daily requirement of calcium. At the $\alpha = 0.1$ level of significance, is there evidence to conclude that the percentage consuming the min daily requirement has increased?

- Step 1: Hypothesis
  H_0:

  H_a:

- Step 2: Conditions
  SRS:

  Independence:

  Normality:

- Step 3: Calculations

  Test Statistic:

- Step 4: Interpretation

Homework: Problems 35, 39, 43, 51

Section 9.3: Tests about a Population Mean

Objectives: Students will:

State and check the Random, 10% and Normal/Large Counts conditions for performing a significance test about a population mean.

Calculate the standardized test statistic and P-value for a test about a population mean

Perform a significance test about a population mean

Use a confidence interval to make a conclusion for a two-sided test about a population

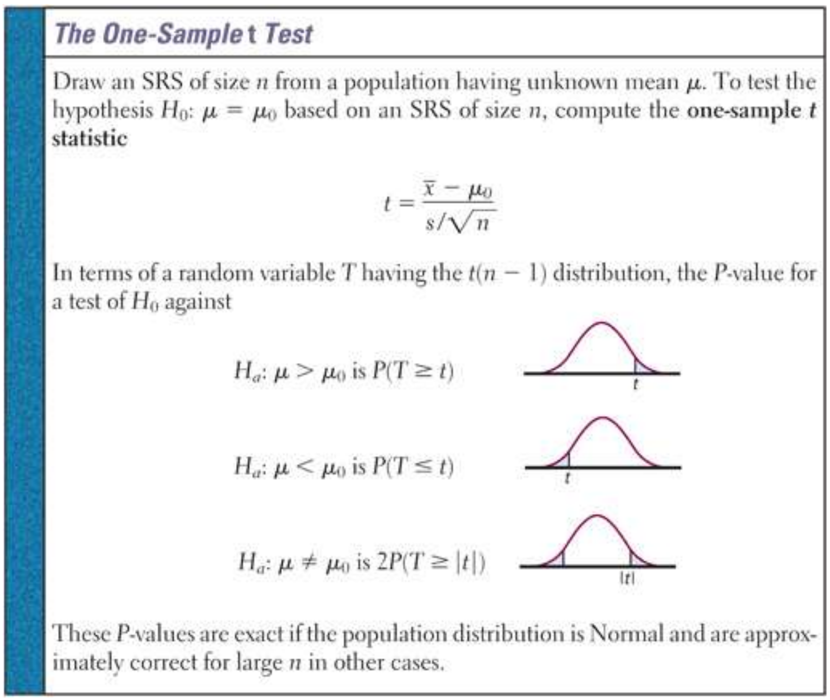
Interpret the power of a significance test and describe what factors affect the power of a test

Vocabulary:

t-distribution – is described by a symmetric, single-peaked, bell-shaped density curve. Any t-distribution is completely specified by its degrees of freedom (df). When performing inference about a population mean based on a random sample of size n when the population standard deviation σ is unknown, use a t-distribution with df = n – 1

power – the probability that the test will find convincing evidence for Ha when a specific alternative value of the parameter is true; also found by power = 1 – β (probability of type II error)

Key Concepts:



The One-Sample t Test

Draw an SRS of size n from a population having unknown mean μ. To test the hypothesis H_0: $\mu = \mu_0$ based on an SRS of size n, compute the **one-sample t statistic**

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

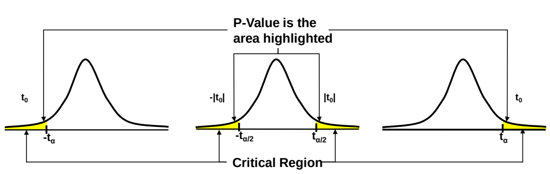
In terms of a random variable T having the $t(n-1)$ distribution, the P-value for a test of H_0 against

H_a: $\mu > \mu_0$ is $P(T \geq t)$

H_a: $\mu < \mu_0$ is $P(T \leq t)$

H_a: $\mu \neq \mu_0$ is $2P(T \geq |t|)$

These P-values are exact if the population distribution is Normal and are approximately correct for large n in other cases.



Test Statistic: $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Reject null hypothesis, if		
P-value < α		
Left-Tailed	Two-Tailed	Right-Tailed
$t_0 < -t_\alpha$	$t_0 < -t_{\alpha/2}$ or $t_0 > t_{\alpha/2}$	$t_0 > t_\alpha$

Confidence Interval Approach

Confidence Interval:

$\bar{x} - t_{\alpha/2} \cdot s/\sqrt{n}$ $\quad$ $\bar{x} + t_{\alpha/2} \cdot s/\sqrt{n}$

Reject null hypothesis, if μ_0 is not in the confidence interval

P-value associated with lower bound must be doubled!

Chapter 9: Testing a Claim

Example 1: Diet colas use artificial sweeteners to avoid sugar. These sweeteners gradually lose their sweetness over time. Trained tasters sip the cola along with drinks of standard sweetness and score the cola on a "sweetness scale" of 1 to 10. The data below is the difference after 4 months of storage in the taster's scores. The bigger these differences, the bigger the loss of sweetness while in storage. Negative values are "gains" in sweetness.

2.0 0.4 0.7 2.0 -0.4 2.2 -1.3 1.2 1.1 2.3

Are these data good evidence that the cola lost sweetness in storage?

- Step 1: Hypothesis
  H_0:

  H_a:

- Step 2: Conditions
  SRS:

  Independence:

  Normality:

- Step 3: Calculations

  Test Statistic:

- Step 4: Interpretation

Example 2: A simple random sample of 12 cell phone bills finds x-bar = $65.014 and s= $18.49. The mean in 2004 was $50.64. Test if the average bill is different today at the $\alpha = 0.05$ level.

- Step 1: Hypothesis
  H_0:

  H_a:

- Step 2: Conditions
  SRS:

  Independence:

  Normality:

- Step 3: Calculations

  Test Statistic:

- Step 4: Interpretation

Example 3: A simple random sample of 40 stay-at-home women finds they watch TV an average of 16.8 hours/week with *s* = 4.7 hours/week. The mean in 2004 was 18.1 hours/week. Test if the average is different today at $\alpha = 0.05$ level.

- Step 1: Hypothesis
  H_0:

  H_a:

- Step 2: Conditions
  SRS:

  Independence:

  Normality:

- Step 3: Calculations

  Test Statistic:

- Step 4: Interpretation

Chapter 9: Testing a Claim

Example 4: To test if pleasant odors improve student performance on tests, 21 subjects worked a paper-and-pencil maze while wearing a mask. The mask was either unscented or carried a floral scent. The response variable is their average time on three trials. Each subject worked the maze with both masks, in a random order (since they tended to improve their times as they worked a maze repeatedly). Assess whether the floral scent significantly improved performance.

Subject	Unscented	Scented		Subject	Unscented	Scented	
1	30.60	37.97		12	58.93	83.50	
2	48.43	51.57		13	54.47	38.30	
3	60.77	56.67		14	43.53	51.37	
4	36.07	40.47		15	37.93	29.33	
5	68.47	49.00		16	43.50	54.27	
6	32.43	43.23		17	87.70	62.73	
7	43.70	44.57		18	53.53	58.00	
8	37.10	28.40		19	64.30	52.40	
9	31.17	28.23		20	47.37	53.63	
10	51.23	68.47		21	53.67	47.00	
11	65.40	51.10					

- Step 1: Hypothesis
  H_0:

  H_a:

- Step 2: Conditions
  SRS:

  Independence:

  Normality:

- Step 3: Calculations

  Test Statistic:

- Step 4: Interpretation

Homework: Problems 65, 69, 73, 81, 85, 87

Chapter 9: Review

Objectives: Students will be able to:

Summarize the chapter
Define the vocabulary used
Know and be able to discuss all sectional knowledge objectives
Complete all sectional construction objectives
Successfully answer any of the review exercises

Conduct *one-sample* and *paired data t significance tests*.
Explain the differences between the one-sample confidence interval for a population proportion and the one-sample significance test for a population proportion.
Conduct a *significance test for a population proportion*.

Vocabulary: None new

TI-83 Calculator Help:

t-Test:

- Press STAT
  - Tab over to TESTS
  - Select T-Test and ENTER
    - Highlight Stats or if Data (id the list its in)
    - Entry μ_0, x-bar, st-dev, and n from summary stats
    - Highlight test type (two-sided, left, or right)
    - Highlight Calculate and ENTER

    Read t-critical and p-value off screen

t-Confidence Interval:

- Press STAT
  - Tab over to TESTS
  - Select Z-Interval and ENTER
    - Highlight Stats
    - Entry s, x-bar, and n from summary stats
    - Entry your confidence level (1- α)
    - Highlight Calculate and ENTER
- Read confidence interval off of screen
  - If μ_0 is in the interval, then FTR
  - If μ_0 is outside the interval, then REJ

One-Sample Proportion Test

- Press STAT
  - Tab over to TESTS
  - Select 1-PropZTest and ENTER
    - Entry p_0, x, and n from given data
    - Highlight test type (two-sided, left, or right)
    - Highlight Calculate and ENTER
- Read z-critical and p-value off screen

Homework: pg775 – 77; 12.31 to 12.38

Chapter 9: Testing a Claim

5 Minute Reviews:

Section 9-1a:

1. The null Hypothesis is

2. The alternative Hypothesis is

3. What different types of alternative Hypotheses do we have?

4. What conditions do we check for before a significance test?

5. What is the p-value?

6. Which types of p-values provide evidence against null hypothesis?

Section 9-1b:

1. What three approaches do we have to inference testing and give their logic?

2. Which ones can be done on our calculator?

3. Can results be statistically significant, but not worth much?

4. What are the two errors that can be done in inference testing and explain when they occur?

Section 9-2:

1. With proportions, when do we reject the null hypothesis?

2. If our normality conditions fail, can we still calculate a p-value? How?

3. What does the confidence interval approach provide that the p-value approach does not?

Chapter 9: Testing a Claim

Section 9-3a:

1. What advantage does using our calculator have over using the t-tables to work the inference test?

2. In a two-sided test, what do we do with alpha (α)?

3. If they do not list an alpha value in the problem, what do we do?

4. If CLT does not apply, what do we do?

5. What do we always watch out for when using a t-distribution?

Section 9-3b:

1. How do we use a confidence interval on a test of population mean?

2. In a matched pair design, what is the data that we use?

3. What is the "power" of an inference test?

4. What type of error is associated with the power of a test?