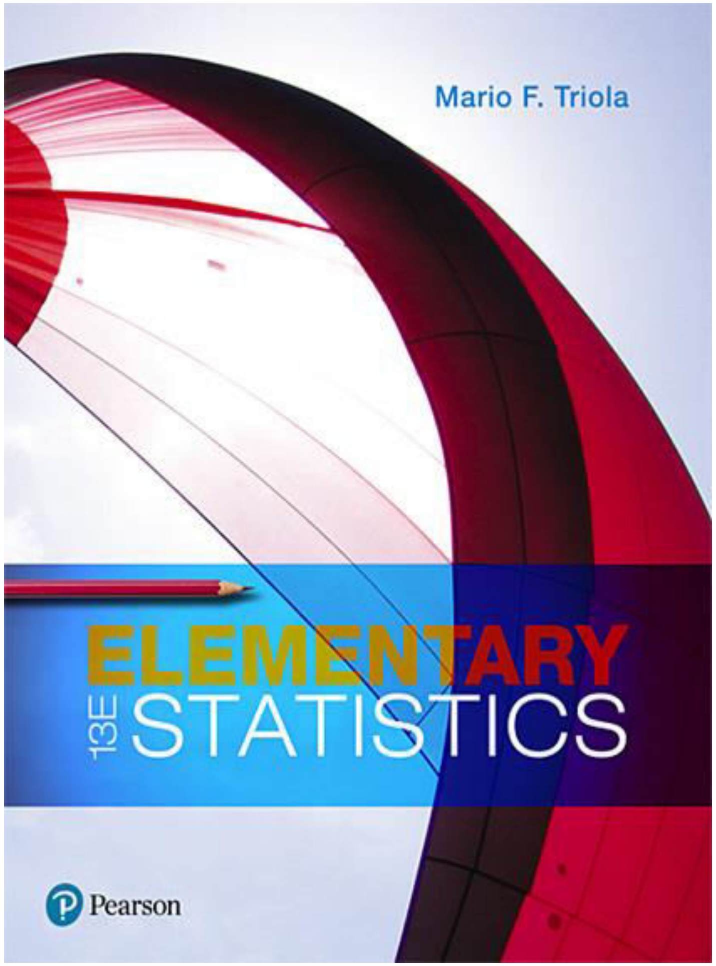
Elementary Statistics

Thirteenth Edition



Chapter 9

Inferences from Two Samples

Inferences from Two Samples

9-1 Two Proportions

9-2 Two Means: Independent Samples

9-3 Two Dependent Samples (Matched Pairs)

9-4 Two Variances or Standard Deviations

Key Concept

In this section we present methods for (1) testing a claim made about two population proportions and (2) constructing a confidence interval estimate of the difference between two population proportions. The methods of this chapter can also be used with probabilities or the decimal equivalents of percentages.

Inferences About Two Proportions: Objectives

Objectives

1. **Hypothesis Test:** Conduct a hypothesis test of a claim about two population proportions.
2. **Confidence Interval:** Construct a confidence interval estimate of the difference between two population proportions.

Inferences About Two Proportions: Notation for Two Proportions

For population 1 we let

p_1 = **population** proportion

$\hat{p}_1 = \dfrac{x_1}{n_1}$ (**sample** proportion)

n_1 = size of the first sample

$\hat{q}_1 = 1 - \hat{p}_1$ (complement of $\hat{p}_1$)

x_1 = number of successes in the first sample

The corresponding notations p_2, n_2, x_2, $\hat{p}_2$, and $\hat{q}_2$ apply to population 2.

Inferences About Two Proportions: Pooled Sample Proportion

The **pooled sample proportion** is denoted by $\bar{p}$ and it combines the two sample proportions into one proportion, as shown here:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{p}$$

Inferences About Two Proportions: Requirements

1. The sample proportions are from two simple random samples.
2. The two samples are **independent.** (Samples are **independent** if the sample values selected from one population are not related to or somehow naturally paired or matched with the sample values from the other population.)
3. For each of the two samples, there are at least 5 successes and at least 5 failures. (That is, $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$ for each of the two samples).

Inferences About Two Proportions: Test Statistic for Two Proportions (with H_0: $p_1 = p_2$)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}}$$

where $p_1 - p_2 = 0$ (assumed in the null hypothesis)

Where $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ (**pooled** sample proportion) and $\bar{q} = 1 - \bar{p}$

P-Value and Critical Values

- ***P*-Value:** *P*-values are automatically provided by technology. If technology is not available, use Table A-2 (standard normal distribution) and find the *P*-value using the procedure given in Figure 8-3 on page 364.
- **Critical Values:** Use Table A-2. (Based on the significance level α, find critical values by using the same procedures introduced in Section 8-1.)

Confidence Interval Estimate of $p_1 - p_2 = 0$

The confidence interval estimate of the difference $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where the margin of error E **is given by**

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}.$$

Hypothesis Tests

For tests of hypotheses made about two population proportions, we consider only tests having a null hypothesis of $p_1 = p_2$ (so the null hypothesis is $H_0 : p_1 = p_2$).

With the assumption that $p_1 = p_2$, the estimates of $\hat{p}_1$ and $\hat{p}_2$ are combined to provide the best estimate of the common value of $\hat{p}_1$ and $\hat{p}_2$, and that combined value is the pooled sample proportion $\bar{p}$ given in the preceding slides.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (1 of 14)

Connecticut and New York are contiguous states, both having laws that require front and rear license plates. The proportion of Connecticut "illegal" cars with rear license plates only is $\frac{239}{2049}$, or 11.7%. The proportion of New York "illegal" cars with rear license plates only is $\frac{9}{550}$, or 1.6%. The sample percentages of 11.7% and 1.6% are obviously different, but are they **significantly** different?

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (2 of 14)

	Connecticut	**New York**
Cars with rear license plate only	239	9
Cars with front and rear license plates	1810	541
Total	**2049**	**550**

$$\text{Connecticut: } \hat{p}_1 = \frac{239}{2049} = 0.117$$

$$\text{New York: } \hat{p}_2 = \frac{9}{550} = 0.016$$

Use a 0.05 significance level and the *P*-value method to test the claim that Connecticut and New York have the same proportion of cars with rear license plates only.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (3 of 14)

Solution

Requirement Check (1) The two samples are simple random samples (trust the author!). (2) The two samples are independent because cars in the samples are not matched or paired in any way. (3) Let’s consider a “success” to be a car with a rear license plate only. For Connecticut, the number of successes is 239 and the number of failures (cars with front and rear license plates) is 1810, so they are both at least 5. For New York, there are 9 successes and 541 failures, and they are both at least 5. The requirements are satisfied.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (4 of 14)

Solution

Step 1: The claim that "Connecticut and New York have the same proportion of cars with rear license plates only" can be expressed as $p_1 = p_2$.

Step 2: If $p_1 = p_2$ is false, then $p_1 \neq p_2$.

Step 3: Because the claim of $p_1 \neq p_2$ does not contain equality, it becomes the alternative hypothesis. The null hypothesis is the statement of equality, so we have

$$H_0: p_1 = p_2 \qquad H_1: p_1 \neq p_2$$

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (5 of 14)

Solution

Step 4: The significance level was specified as $\alpha = 0.05$, so we use $\alpha = 0.05$.

Step 5: This step and the following step can be circumvented by using technology; see the display that follows this example. If not using technology, we use the normal distribution as an approximation to the binomial distribution.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (6 of 14)

Solution

Step 5 (con't): We estimate the common value of p_1 and p_2 with the pooled sample estimate p calculated as shown below, with extra decimal places used to minimize rounding errors in later calculations.

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{239 + 9}{2049 + 550} = 0.09542132$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.09542132 = 0.90457868.$$

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (7 of 14)

Solution

Step 6: Because we assume in the null hypothesis that $p_1 = p_2$, the value of $p_1 - p_2$ is 0 in the following calculation of the test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$$= \frac{\left(\frac{239}{2049} - \frac{9}{550}\right) - 0}{\sqrt{\frac{(0.09542132)(0.90457868)}{2049} + \frac{(0.09542132)(0.90457868)}{550}}}$$

$$= 7.11$$

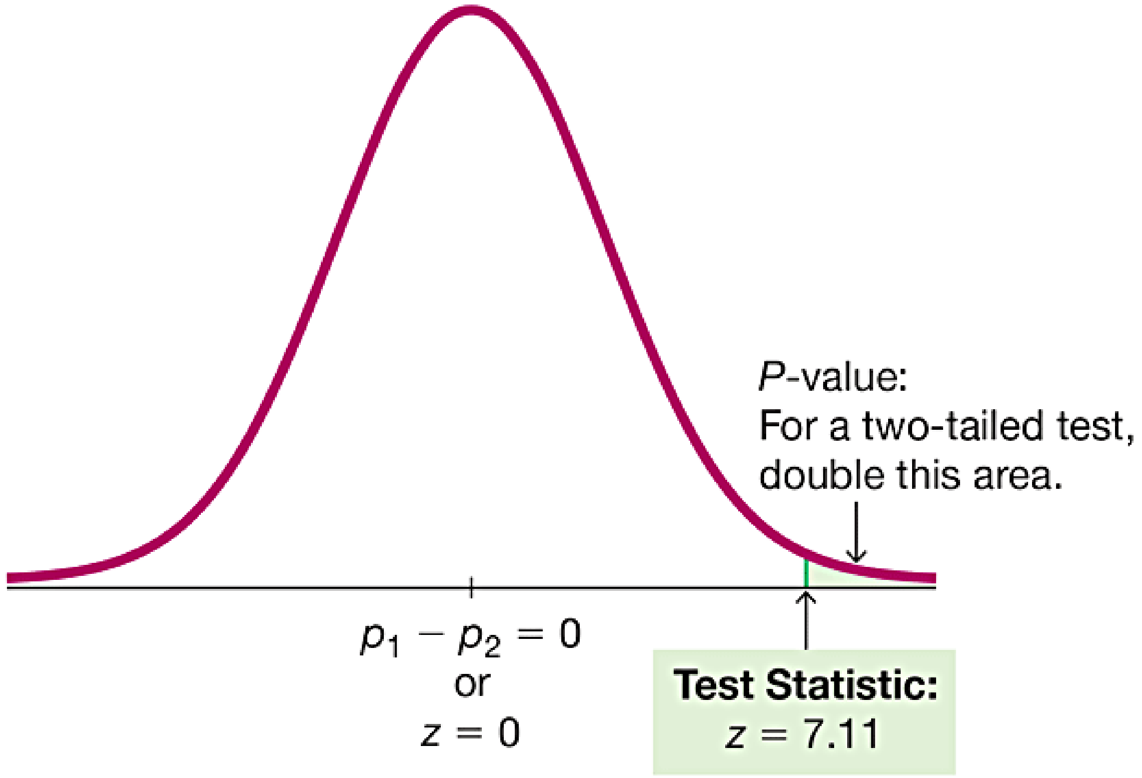
Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (8 of 14)

Solution

Step 6 (con't): This is a two-tailed test, so the *P*-value is twice the area to the right of the test statistic $z = 7.11$. Refer to Table A-2 and find that the area to the right of the test statistic $z = 7.11$ is 0.0001, so the *P*-value is 0.0002. Technology provides a more accurate *P*-value of 0.00000000000119, which is often expressed as 0.0000 or "*P*-value < 0.0001."

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (9 of 14)

Solution



Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (10 of 14)

Solution

Step 7: Because the *P*-value of 0.0000 is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis of $p_1 = p_2$.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (11 of 14)

Interpretation

We must address the original claim that “Connecticut and New York have the same proportion of cars with rear license plates only.” Because we reject the null hypothesis, we conclude that there is sufficient evidence to warrant rejection of the claim that $p_1 = p_2$. That is, there is sufficient evidence to conclude that Connecticut and New York have **different** proportions of cars with rear license plates only. It’s reasonable to speculate that enforcement of the license plate laws is much stricter in New York than in Connecticut, and that is why Connecticut car owners are less likely to install the front license plate.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (12 of 14)

Technology

Software and calculators usually provide a *P*-value, so the *P*-value method is typically used for testing a claim about two proportions. See the Statdisk results showing the test statistic of $z = 7.11$(rounded) and the *P*-value of 0.0000.

Statdisk

Pooled proportion: 0.0954213
Test Statistic, z: 7.1074
Critical z: ±1.9600
P-Value: 0.0000

95% Confidence interval:
0.0827974 < p1-p2 < 0.1177599

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (13 of 14)

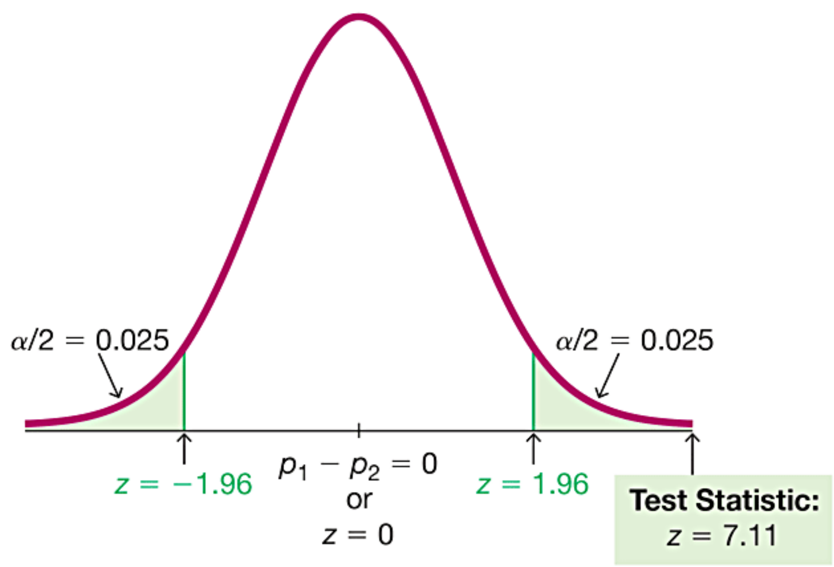
Solution (Critical Value Method)

The critical value method of testing hypotheses can also be used. In Step 6, find the critical values. With a significance level of $\alpha = 0.05$ in a two-tailed test based on the normal distribution, we refer to Table A-2 and find that an area of $\alpha = 0.05$ divided equally between the two tails corresponds to the critical values of $z = \pm 1.96$.

Example: Proportions of Cars with Rear License Plates Only: Are the Proportions the Same in Connecticut and New York? (14 of 14)

Solution (Critical Value Method)

We can see that the test statistic of z = 7.11 falls within the critical region beyond the critical value of 1.96. We again reject the null hypothesis. The conclusions are the same.



Pearson

Example: Confidence Interval for Claim About Two Proportions (1 of 5)

Use the same sample data given to construct a 95% confidence interval estimate of the difference between the two population proportions. What does the result suggest about the claim that “Connecticut and New York have the same proportion of cars with rear license plates only”?

Example: Confidence Interval for Claim About Two Proportions (2 of 5)

Solution

Requirement Check We are using the same data and the same requirement check applies here. The confidence interval can be found using technology.

Statdisk

Pooled proportion: 0.0954213
Test Statistic, z: 7.1074
Critical z: ±1.9600
P-Value: 0.0000

95% Confidence interval:
0.0827974 < p1-p2 < 0.1177599

Example: Confidence Interval for Claim About Two Proportions (3 of 5)

Solution

If not using technology, proceed as follows.

With a 95% confidence level, $z_{\frac{\alpha}{2}} = 1.96$. We calculate the value of the margin of error E as shown here.

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$= 1.96 \sqrt{\frac{\left(\frac{239}{2049}\right)\left(\frac{1810}{2049}\right)}{2049} + \frac{\left(\frac{9}{550}\right)\left(\frac{541}{550}\right)}{550}} = 0.017482$$

Example: Confidence Interval for Claim About Two Proportions (4 of 5)

Solution

With $\hat{p}_1 = \dfrac{239}{2049} = 0.116642$ and $\hat{p}_2 = \dfrac{9}{550} = 0.016364$, we get $\hat{p}_1 - \hat{p}_2 = 0.100278$. With $E = 0.017482$, the confidence interval is evaluated as follows, with the confidence interval limits rounded to three significant digits:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$0.100278 - 0.017482 < (p_1 - p_2) < 0.100278 + 0.017482$$

$$0.0828 < (p_1 - p_2) < 0.118$$

See the preceding Statdisk display showing the same confidence interval obtained here.

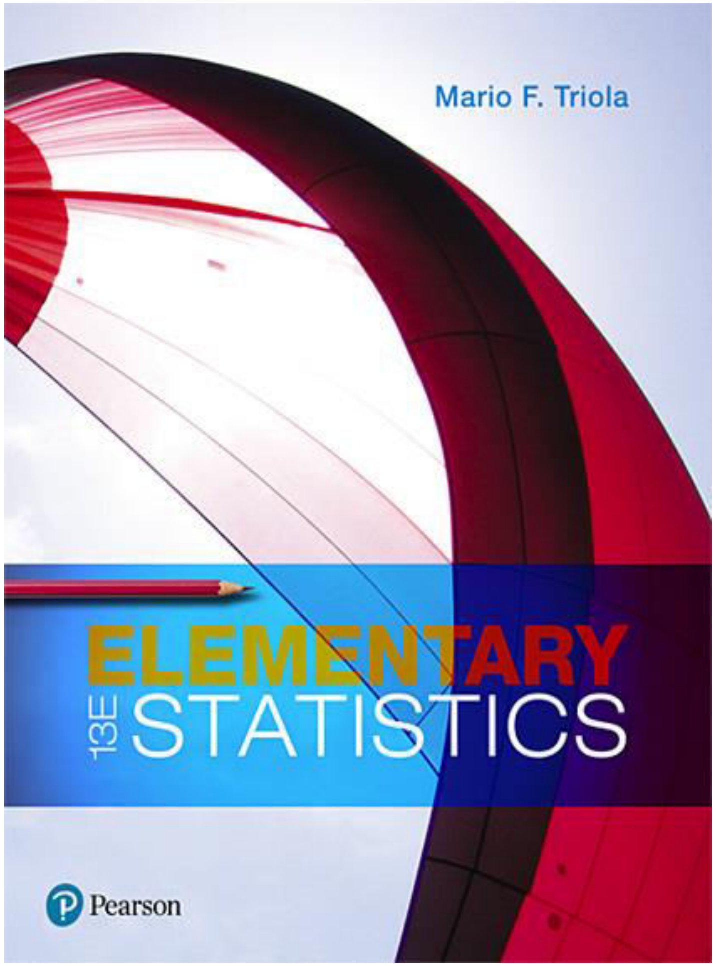
Example: Confidence Interval for Claim About Two Proportions (5 of 5)

Interpretation

The confidence interval limits do not contain 0, suggesting that there is a significant difference between the two proportions. The confidence interval suggests that the value of p_1 is greater than the value of p_2, so there does appear to be sufficient evidence to warrant rejection of the claim that “Connecticut and New York have the same proportion of cars with rear license plates only.”

Elementary Statistics

Thirteenth Edition



Chapter 9

Inferences from Two Samples

Inferences from Two Samples

9-1 Two Proportions

9-2 Two Means: Independent Samples

9-3 Two Dependent Samples (Matched Pairs)

9-4 Two Variances or Standard Deviations

Key Concept

In this section we present the *F* test for testing claims made about two population variances (or standard deviations). The *F* test (named for statistician Sir Ronald Fisher) uses the *F* distribution introduced in this section. The *F* test requires that both populations have normal distributions. Instead of being robust, this test is **very** sensitive to departures from normal distributions, so the normality requirement is quite strict.

Hypothesis Test with Two Variances or Standard Deviations: Objective

Conduct a hypothesis test of a claim about two population variances or standard deviations. (Any claim made about two population standard deviations can be restated with an equivalent claim about two population variances, so the same procedure is used for two population standard deviations or two population variances.)

Hypothesis Test with Two Variances or Standard Deviations: Notation

s_1^2 = **larger** of the two sample variances

n_1 = size of the sample with the **larger** variance

σ_1^2 = variance of the population from which the sample with the **larger** variance was drawn

The symbols s_2^2, n_2, and σ_2^2 are used for the other sample and population.

Hypothesis Test with Two Variances or Standard Deviations: Requirements

1. The two populations are **independent.**
2. The two samples are simple random samples.
3. Each of the two populations must be **normally distributed**, regardless of their sample sizes. This F test is **not robust** against departures from normality, so it performs poorly if one or both of the populations have a distribution that is not normal. The requirement of normal distributions is quite strict for this F test.

Hypothesis Test with Two Variances or Standard Deviations: Test Statistic for Hypothesis Tests with Two Variances (with H_0: sigma 1 squared equal to sigma 2 squared) (1 of 4)

$$F = \frac{s_1^2}{s_2^2}$$

(where s_1^2 is the **larger** of the two sample variances)

***P*-Values:** *P*-values are automatically provided by technology. If technology is not available, use the computed value of the *F* test statistic with Table A-5 to find a range for the *P*-value.

Hypothesis Test with Two Variances or Standard Deviations: Test Statistic for Hypothesis Tests with Two Variances (with H_0: sigma 1 squared equal to sigma 2 squared) (2 of 4)

Critical Values: Use Table A-5 to find critical *F* values that are determined by the following:

1. The significance level α (Table A-5 includes critical values for $\alpha = 0.025$ and $\alpha = 0.05$.)
2. **Numerator degrees of freedom $n_1 - 1$** (determines **column** of Table A-5)

Pearson

Hypothesis Test with Two Variances or Standard Deviations: Test Statistic for Hypothesis Tests with Two Variances (with H_0: sigma 1 squared equal to sigma 2 squared) (3 of 4)

3. **Denominator degrees of freedom $n_2 - 1$** (determines **row** of Table A-5) For significance level $\alpha = 0.05$, refer to Table A-5 and use the right-tail area of 0.025 or 0.05, depending on the type of test, as shown below:
   - **Two-tailed test:** Use Table A-5 with 0.025 in the right tail. (The significance level of 0.05 is divided between the two tails, so the area in the right tail is 0.025.)
   - **One-tailed test:** Use Table A-5 with $\alpha = 0.05$ in the right tail.

Pearson

Hypothesis Test with Two Variances or Standard Deviations: Test Statistic for Hypothesis Tests with Two Variances (with H_0: sigma 1 squared equal to sigma 2 squared) (4 of 4)

Find the critical *F* value for the right tail: Because we are stipulating that the larger sample variance is s_1^2, all one-tailed tests will be right-tailed and all two-tailed tests will require that we find only the critical value located to the right. (We have no need to find the critical value at the left tail, which is not very difficult.)

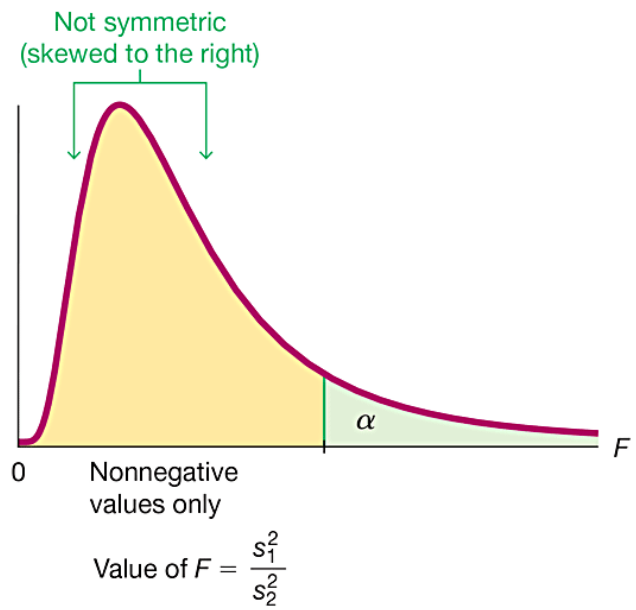
F Distribution (1 of 2)

- *F* Distribution
  - For two normally distributed populations with equal variances $(\sigma_1^2 = \sigma_2^2)$, the sampling distribution of the test statistic $F = \dfrac{s_1^2}{s_2^2}$ is the ***F* distribution.**

F Distribution (2 of 2)

There is a different *F* distribution for each different pair of degrees of freedom for the numerator and denominator.

- The *F* distribution is not symmetric.
- Values of the *F* distribution cannot be negative.
- The exact shape of the *F* distribution depends on the two different degrees of freedom.



Interpreting the Value of the *F* Test Statistic

If the two populations have equal variances, then the ratio $\frac{s_1^2}{s_2^2}$ will tend to be close to 1. Because we are stipulating that s_1^2 is the larger sample variance, the ratio $\frac{s_1^2}{s_2^2}$ will be a **large** number whenever s_1^2 and s_2^2 are far apart in value. Consequently, a value of *F* near 1 will be evidence in favor of $\sigma_1^2 = \sigma_2^2$, but a large value of *F* will be evidence against $\sigma_1^2 = \sigma_2^2$.

Large value of *F* will be evidence against $\sigma_1^2 = \sigma_2^2$.

Example: Course Evaluation Scores (1 of 9)

Listed below are the same student course evaluation scores used in Section 9-2, where we tested the claim that the two samples are from populations with the same **mean.** Use the same data with a 0.05 significance level to test the claim that course evaluation scores of female professors and male professors have the same **variation.**

Female	4.3	4.3	4.4	4.0	3.4	4.7	2.9	4.0	4.3	3.4	3.4	3.3			
Male	4.5	3.7	4.2	3.9	3.1	4.0	3.8	3.4	4.5	3.8	4.3	4.4	4.1	4.2	4.0

Example: Course Evaluation Scores (2 of 9)

Solution

REQUIREMENT CHECK (1) The two populations are independent of each other. The two samples are not matched in any way. (2) Given the design for the study, we assume that the two samples can be treated as simple random samples. (3) A normal quantile plot of each set of sample course evaluation scores shows that both samples appear to be from populations with a normal distribution.

The requirements are satisfied.

Example: Course Evaluation Scores (3 of 9)

Solution

For females, we get $s = 0.5630006$ and for males we get $s = 0.3954503$. We can conduct the test using either variances or standard deviations. Because we stipulate in this section that the larger variance is denoted by s_1^2, we let $s_1^2 = 0.5630006^2$ and $s_2^2 = 0.3954503^2$.

Example: Course Evaluation Scores (4 of 9)

Solution

Step 1: The claim that male and female professors have the same variation can be expressed symbolically as $\sigma_1^2 = \sigma_2^2$ or as $\sigma_1 = \sigma_2$. Use $\sigma_1 = \sigma_2$.

Step 2: If the original claim is false, then $\sigma_1 \neq \sigma_2$.

Step 3: Because the null hypothesis is the statement of equality and because the alternative hypothesis cannot contain equality, we have

$$H_0 : \sigma_1 = \sigma_2 \text{(original claim)} \qquad H_1 : \sigma_1 \neq \sigma_2$$

Example: Course Evaluation Scores (5 of 9)

Solution

Step 4: The significance level is $\alpha = 0.05$.

Step 5: Because this test involves two population variances, we use the F distribution.

Step 6: The test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{0.5630006^2}{0.3954503^2} = 2.0269$$

Example: Course Evaluation Scores (6 of 9)

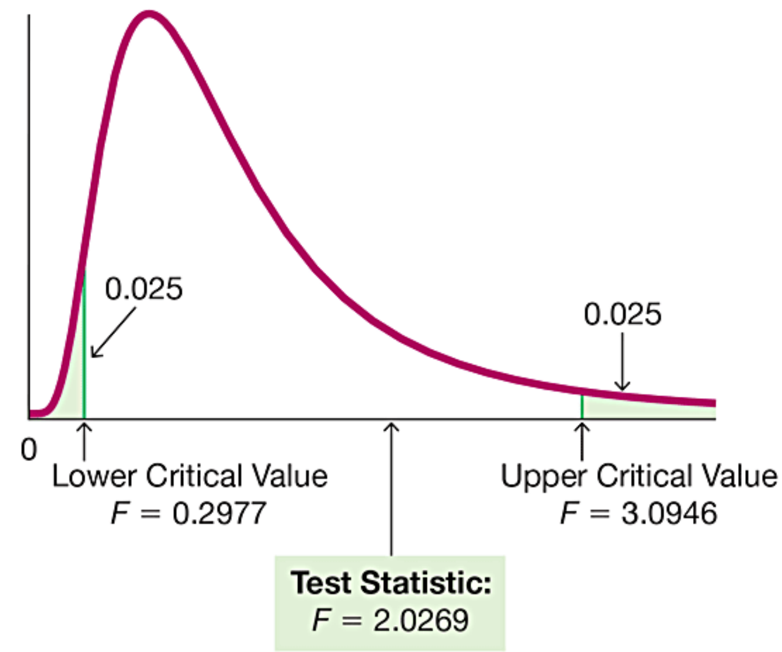
Solution

***P*-Value Method** For a two-tailed test with significance level 0.05, the area of 0.025 is in the right tail. With numerator degrees of freedom = $n_1 - 1 = 11$ and denominator degrees of freedom = $n_2 - 1 = 14$, we find that the critical value of F is between 3.1469 and 3.0502. The test statistic of $F = 2.0269$ is less than the critical value, so the area to the right of the test statistic is greater than 0.025, and for this two-tailed test, P-value > 0.05.

Example: Course Evaluation Scores (7 of 9)

Solution

Critical Value Method As with the *P*-value method, we find that the critical value is between 3.1469 and 3.0502. The test statistic $F = 2.0269$ is less than the critical value (which is between 3.1469 and 3.0502), so the test statistic does not fall in the critical region.



Example: Course Evaluation Scores (8 of 9)

Solution

Step 7: The test statistic $F = 2.0269$ does not fall within the critical region, so we fail to reject the null hypothesis of equal variances. There is not sufficient evidence to warrant rejection of the claim of equal standard deviations.

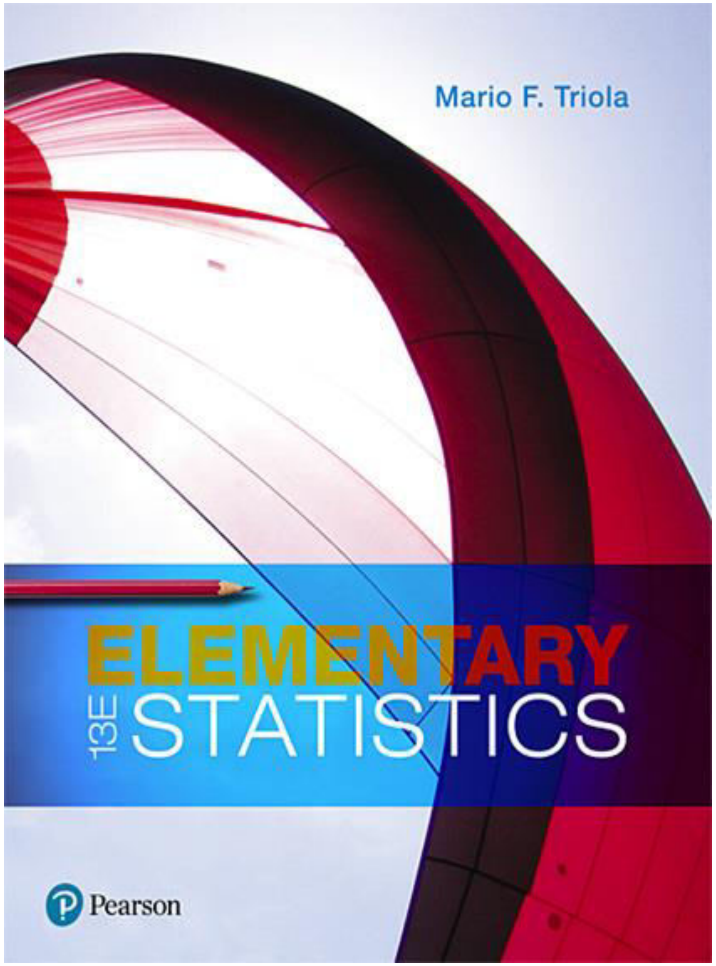
Example: Course Evaluation Scores (9 of 9)

Interpretation

Step 8: There is not sufficient evidence to warrant rejection of the claim that the two standard deviations are equal. Course evaluation scores of female professors and male professors appear to have the same amount of variation.

Elementary Statistics

Thirteenth Edition



Chapter 9

Inferences from Two Samples

Inferences from Two Samples

9-1 Two Proportions

9-2 Two Means: Independent Samples

9-3 Two Dependent Samples (Matched Pairs)

9-4 Two Variances or Standard Deviations

Key Concept

This section presents methods for testing hypotheses and constructing confidence intervals involving the mean of the differences of the values from two populations that are dependent in the sense that the data consist of matched pairs. The pairs must be matched according to some relationship, such as before/after measurements from the same subjects or IQ scores of husbands and wives.

Good Experimental Design

When designing an experiment or planning an observational study, using dependent samples with matched pairs is generally better than using two independent samples.

Inferences About Differences from Matched Pairs: Objectives

1. **Hypothesis Test:** Use the differences from two dependent samples (matched pairs) to test a claim about the mean of the population of all such differences.
2. **Confidence Interval:** Use the differences from two dependent samples (matched pairs) to construct a confidence interval estimate of the mean of the population of all such differences.

Pearson

Inferences About Differences from Matched Pairs: Notation for Dependent Samples

- d = individual difference between the two values in a single matched pair
- μ_d = mean value of the differences d for the **population** of all matched pairs of data
- $\bar{d}$ = mean value of the differences d for the paired **sample** data
- s_d = standard deviation of the differences d for the paired **sample** data
- n = number of **pairs** of sample data

Inferences About Differences from Matched Pairs: Requirements

1. The sample data are dependent (matched pairs).
2. The matched pairs are a simple random sample.
3. Either or both of these conditions are satisfied: The number of pairs of sample data is large ($n > 30$) or the pairs of values have differences that are from a population having a distribution that is approximately normal. These methods are **robust** against departures for normality, so the normality requirement is loose.

Inferences About Differences from Matched Pairs: Test Statistic for Dependent Samples (with H_0: $\mu_d = 0$) (1 of 2)

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

***P*-Values:** *P*-values are automatically provided by technology or the *t* distribution in Table A-3 can be used.

Inferences About Differences from Matched Pairs: Test Statistic for Dependent Samples (with H_0: $\mu_d = 0$) (2 of 2)

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

Critical Values: Use Table A-3 (*t* distribution). For degrees of freedom, use df = $n - 1$.

Inferences About Differences from Matched Pairs: Confidence Intervals for Dependent Samples

$$\bar{d} - E < \mu_d < \bar{d} + E$$

where $E = t_{\frac{\alpha}{2}} \dfrac{s_d}{\sqrt{n}}$ (Degrees of freedom: df = n – 1.)

Procedures for Inferences with Dependent Samples

1. Verify that the sample data consist of dependent samples (or matched pairs), and verify that the requirements in the preceding slides are satisfied.
2. Find the difference d for each pair of sample values.
3. Find the value of $\bar{d}$ and s_d.
4. For hypothesis tests and confidence intervals, use the same t test procedures used for a single population mean.

Equivalent Methods

Because the hypothesis test and confidence interval in this section use the same distribution and standard error, they are **equivalent** in the sense that they result in the same conclusions. Consequently, a null hypothesis that the mean difference equals 0 can be tested by determining whether the confidence interval includes 0.

Example: Are Best Actresses Generally Younger Than Best Actors? (1 of 9)

Data lists ages of actresses when they won Oscars in the category of Best Actress, along with the ages of actors when they won Oscars in the category of Best Actor. The ages are matched according to the year that the awards were presented. This is a small random selection of the available data so that we can better illustrate the procedures of this section. Use the sample data with a 0.05 significance level to test the claim that for the population of ages of Best Actresses and Best Actors, the differences have a mean less than 0.

Actress (years)	28	28	31	29	35
Actor (years)	62	37	36	38	29
Difference *d*	−34	−9	−5	−9	6

Example: Are Best Actresses Generally Younger Than Best Actors? (2 of 9)

Solution

Requirement Check (1) The samples are dependent because the values are matched by the year in which the awards were given. (2) The pairs of data are randomly selected. We will consider the data to be a simple random sample. (3) Because the number of pairs of data is $n = 5$, which is not large, we should check for normality of the differences and we should check for outliers. There are no outliers, and a normal quantile plot would show that the points approximate a straight-line pattern with no other pattern.

All requirements are satisfied.

Example: Are Best Actresses Generally Younger Than Best Actors? (3 of 9)

Solution

Step 1: The claim that the differences have a mean less than 0 can be expressed as $\mu_d < 0$ year.

Step 2: If the original claim is not true, we have $\mu_d \geq 0$ year.

Step 3: The null hypothesis must express equality and the alternative hypothesis cannot include equality, so we have

H_0: $\mu_d = 0$ year H_1: $\mu_d < 0$ year (original claim)

Example: Are Best Actresses Generally Younger Than Best Actors? (4 of 9)

Solution

Step 4: The significance level is $\alpha = 0.05$.

Step 5: We use the Student t distribution.

Step 6: Before finding the value of the test statistic, we must first find the values of $\bar{d}$ and s_d. We use the differences (−34, −9, −5, −9, 6) to find these sample statistics: $\bar{d} = -10.2$ years and $s_d = 14.7$ years.

Example: Are Best Actresses Generally Younger Than Best Actors? (5 of 9)

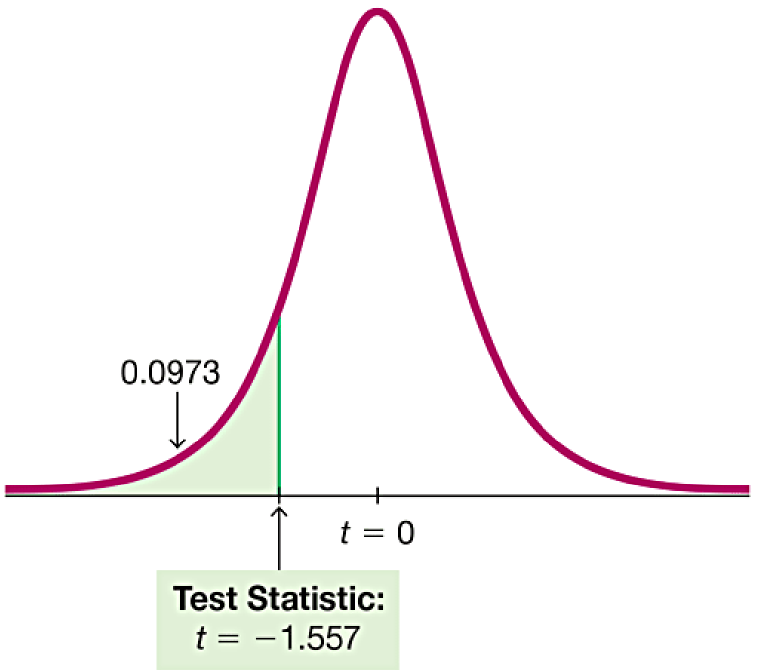
Solution

Step 6 (con't): Using these sample statistics and the assumption from the null hypothesis that $\mu_d = 0$ year, we can now find the value of the test statistic. (The value of $t = -1.557$ is obtained if unrounded values of $\bar{d}$ and s_d are used; technology will provide a test statistic of $t = -1.557$.)

Example: Are Best Actresses Generally Younger Than Best Actors? (6 of 9)

Solution

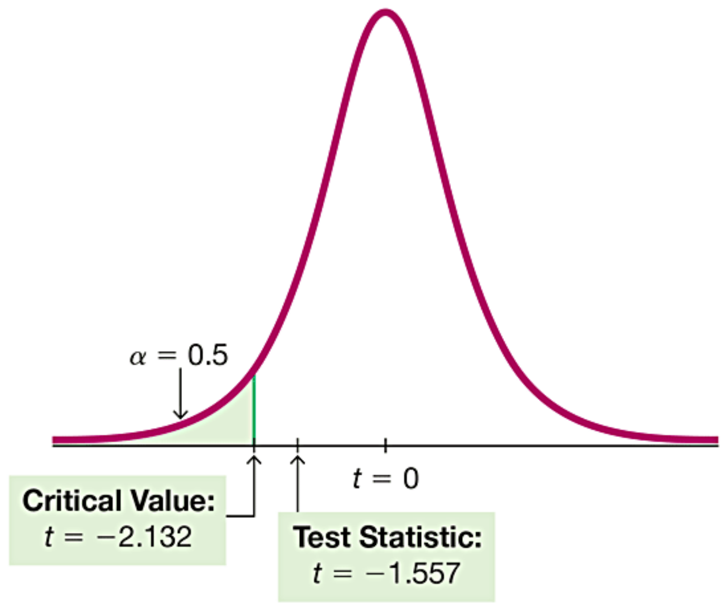
***P*-Value Method** Because we are using a *t* distribution, we refer to Table A-3 for the row with df = 4 and we see that the test statistic $t = -1.552$ corresponds to an "Area in One Tail" that is greater than 0.05, so *P*-value > 0.05. Technology would provide *P*-value = 0.0973.



Example: Are Best Actresses Generally Younger Than Best Actors?

Solution

Critical Value Method Refer to Table A-3 to find the critical value of $t = -2.132$ as follows: Use the column for 0.05 (Area in One Tail), and use the row with degrees of freedom of $n - 1 = 4$. The critical value $t = -2.132$ is negative because this test is left-tailed where all values of t are negative.



Example: Are Best Actresses Generally Younger Than Best Actors?

Solution

Step 7: If we use the *P*-value method, we fail to reject H_0 because the *P*-value is greater than the significance level of 0.05. If we use the critical value method, we fail to reject H_0 because the test statistic does not fall in the critical region.

Example: Are Best Actresses Generally Younger Than Best Actors? (9 of 9)

Interpretation

We conclude that there is not sufficient evidence to support $\mu_d < 0$. There is not sufficient evidence to support the claim that for the population of ages of Best Actresses and Best Actors, the differences have a mean less than 0. There is not sufficient evidence to conclude that Best Actresses are generally younger than Best Actors.

Example: Confidence Interval for Estimating the Mean of the Age Differences (1 of 4)

Using the same sample data, construct a 90% confidence interval estimate of μ_d, which is the mean of the age differences. By using a confidence level of 90%, we get a result that could be used for the previous hypothesis test. (Because the hypothesis test is one-tailed with a significance level of $\alpha = 0.05$, the confidence level should be 90%.)

Example: Confidence Interval for Estimating the Mean of the Age Differences (2 of 4)

Solution

REQUIREMENT CHECK The solution for previous example includes verification that the requirements are satisfied.

The Statdisk display shows the 90% confidence interval. It is found using the values of $\bar{d} = -10.2$ years, $s_d = 14.7$ years, and $t_{\frac{\alpha}{2}} = 2.132$ (found from Table A-3 with $n - 1 = 4$ degrees of freedom and an area of 0.10 divided equally between the two tails).

Statdisk

Sample size, n: 5
Difference Mean, d: -10.2
Difference Standard Deviation, sd: 14.65264
Test Statistic, t: -1.5566
Critical t: -2.1318
P-Value: 0.0973

90% Confidence interval:
-24.16971 < μd < 3.769712

Example: Confidence Interval for Estimating the Mean of the Age Differences (3 of 4)

Solution

We first find the value of the margin of error *E*.

$$E = t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}} = 2.132 \cdot \frac{14.7}{\sqrt{5}} = 14.015853$$

We now find the confidence interval.

$$\bar{d} - E < \mu_d < \bar{d} + E$$

$$-10.2 - 14.015853 < \mu_d < -10.2 + 14.015853$$

$$-24.2 \text{ years} < \mu_d < 3.8 \text{ years}$$

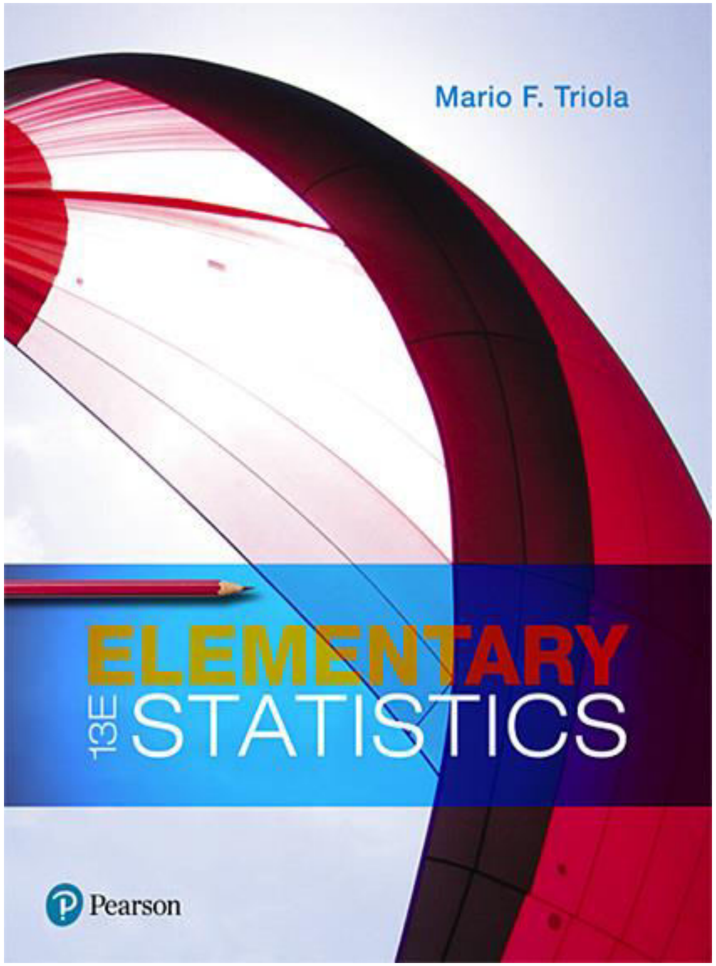
Example: Confidence Interval for Estimating the Mean of the Age Differences

Interpretation

We have 90% confidence that the limits of −24.2 years and 3.8 years contain the true value of the mean of the age differences. In the long run, 90% of such samples will lead to confidence interval limits that actually do contain the true population mean of the differences. The confidence interval includes the value of 0 year, so it is possible that the mean of the differences is equal to 0 year, indicating that there is no significant difference between ages of Best Actresses and Best Actors.

Elementary Statistics

Thirteenth Edition



Chapter 9

Inferences from Two Samples

Inferences from Two Samples

9-1 Two Proportions

9-2 Two Means: Independent Samples

9-3 Two Dependent Samples (Matched Pairs)

9-4 Two Variances or Standard Deviations

Key Concept

This section presents methods for using sample data from two independent samples to test hypotheses made about two population means or to construct confidence interval estimates of the difference between two population means.

Independent (1 of 2)

- Independent
  - Two samples are **independent** if the sample values from one population are not related to or somehow naturally paired or matched with the sample values from the other population.

Independent (2 of 2)

- Dependent
  - Two samples are **dependent** (or consist of **matched pairs**) if the sample values are somehow matched, where the matching is based on some inherent relationship.

Inferences About Two Means: Independent Samples: Objectives

Objectives

1. **Hypothesis Test:** Conduct a hypothesis test of a claim about two independent population means.
2. **Confidence Interval:** Construct a confidence interval estimate of the difference between two independent population means.

Inferences About Two Means: Independent Samples: Notation

Notation

For population 1 we let

μ_1 = **population** mean

$\bar{x}_1$ = **sample** mean

σ_1 = **population** standard deviation

s_1 = **sample** standard deviation

n_1 = size of the first sample

The corresponding notations μ_2, s_2, $\bar{x}_2$, s_2 , and n_2, apply to population 2.

Inferences About Two Means: Independent Samples: Requirements

Requirements

1. The values of σ_1 and σ_2 are unknown and we do not assume that they are equal.
2. The two samples are **independent.**
3. Both samples are **simple random samples**.
4. Either or both of these conditions are satisfied: The two sample sizes are both **large** (with $n_1 > 30$ and $n_2 > 30$) or both samples come from populations having normal distributions.

Inferences About Two Means: Independent Samples: Hypothesis Test Statistic for Two Means: Independent Samples (with H_0: $\mu_1 = \mu_2$) (1 of 3)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

(where $\mu_1 - \mu_2$ is often assumed to be 0)

Pearson

Inferences About Two Means: Independent Samples: Hypothesis Test Statistic for Two Means: Independent Samples (with H_0: $\mu_1 = \mu_2$) (2 of 3)

Degrees of freedom

1. Use this simple and conservative estimate:
   df = smaller of $n_1 - 1$ and $n_2 - 1$.
2. Technologies typically use the more accurate but more difficult estimate given by the following formula:

$$\text{df} = \frac{(A+B)^2}{\dfrac{A^2}{n_1 - 1} + \dfrac{B^2}{n_2 - 1}} \quad \text{where } A = \frac{s_1^2}{n_1} \text{ and } B = \frac{s_2^2}{n_2}$$

Inferences About Two Means: Independent Samples: Hypothesis Test Statistic for Two Means: Independent Samples (with H_0: $\mu_1 = \mu_2$) (3 of 3)

***P*-Values:** *P*-values are automatically provided by technology. If technology is not available, refer to the *t* distribution in Table A-3. **Critical Values:** Refer to the *t* distribution in Table A-3.

Inferences About Two Means: Independent Samples: Confidence Interval Estimate of $\mu_1 - \mu_2$: Independent Samples

The confidence interval estimate of the difference $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

where the margin of error E **is given by**

$$E = t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Example: Are Male Professors and Female Professors Rated Differently by Students? (1 of 10)

Listed below are student course evaluation scores for courses taught by female professors and male professors. Use a 0.05 significance level to test the claim that the two samples are from populations with the same mean. Does there appear to be a difference in evaluation scores of courses taught by female professors and male professors?

Female	4.3	4.3	4.4	4.0	3.4	4.7	2.9	4.0	4.3	3.4	3.4	3.3			
Male	4.5	3.7	4.2	3.9	3.1	4.0	3.8	3.4	4.5	3.8	4.3	4.4	4.1	4.2	4.0

Example: Are Male Professors and Female Professors Rated Differently by Students? (2 of 10)

Solution

Requirement Check (1) The values of the two population standard deviations are not known and we are not making an assumption that they are equal. (2) The two samples are independent because the female professors and male professors are not matched or paired in any way.

Example: Are Male Professors and Female Professors Rated Differently by Students?

Solution

Requirement Check (3) The samples are simple random samples. (4) Both samples are small (30 or fewer), so we need to determine whether both samples come from populations having normal distributions. Normal quantile plots of the two samples suggest that the samples are from populations having distributions that are not far from normal.

The requirements are all satisfied.

Example: Are Male Professors and Female Professors Rated Differently by Students? (4 of 10)

Solution

Step 1: The claim that "the two samples are from populations with the same mean" can be expressed as $\mu_1 = \mu_2$.

Step 2: If the original claim is false, then $\mu_1 \neq \mu_2$.

Step 3: The alternative hypothesis is the expression not containing equality, and the null hypothesis is an expression of equality, so we have

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

Example: Are Male Professors and Female Professors Rated Differently by Students? (5 of 10)

Solution

We now proceed with the assumption that $\mu_1 = \mu_2$, or $\mu_1 - \mu_2 = 0$.

Step 4: The significance level is $\alpha = 0.05$.

Step 5: Because we have two independent samples and we are testing a claim about the two population means, we use a *t* distribution with the test statistic given earlier in this section.

Example: Are Male Professors and Female Professors Rated Differently by Students? (6 of 10)

Solution

Step 6: The test statistic is calculated using these statistics (with extra decimal places) obtained from the listed sample data:

Females: $n = 12$, $\bar{x} = 3.866667$, $s = 0.563001$; males: $n = 15$, $\bar{x} = 3.993333$, $s = 0.395450$.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(3.866667 - 3.993333) - 0}{\sqrt{\frac{0.563001^2}{12} + \frac{0.395450^2}{15}}} = -0.660$$

Example: Are Male Professors and Female Professors Rated Differently by Students? (7 of 10)

Solution

***P*-Value** With test statistic $t = -0.660$, we refer to Table A-3 (t Distribution). The number of degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$, or the smaller of $(12 - 1)$ and $(15 - 1)$, which is 11. With df = 11 and a two-tailed test, Table A-3 indicates that the *P*-value is greater than 0.20. Technology will provide the *P*-value of 0.5172 when using the original data or unrounded sample statistics.

Example: Are Male Professors and Female Professors Rated Differently by Students? (8 of 10)

Solution

Step 7: Because the *P*-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. ("If the *P* is **low,** the null must go.")

Example: Are Male Professors and Female Professors Rated Differently by Students? (9 of 10)

Interpretation

Step 8: There is not sufficient evidence to warrant rejection of the claim that female professors and male professors have the same mean course evaluation score.

Example: Are Male Professors and Female Professors Rated Differently by Students?

Technology

The tricky part about the preceding *P*-value approach is that Table A-3 can give only a range for the *P*-value, and determining that range is often somewhat difficult. Technology automatically provides the *P*-value, so technology makes the *P*-value method quite easy. See the accompanying XLSTAT display showing the test statistic of $t = -0.660$ (rounded) and the *P*-value of 0.5172.

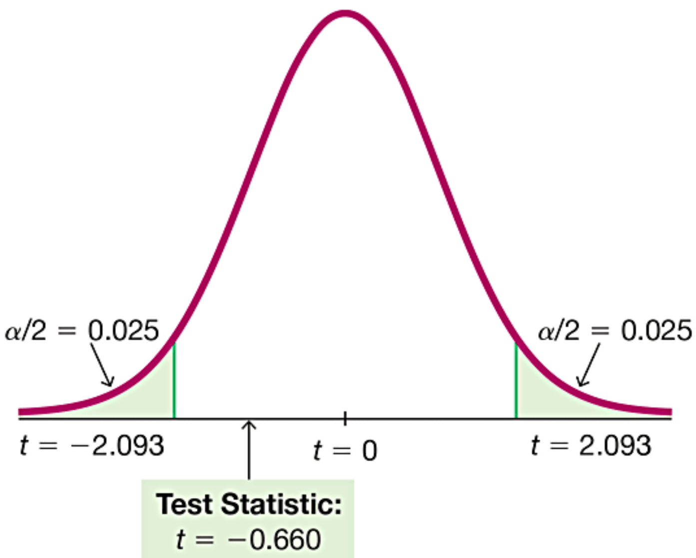
XLSTAT

Difference	-0.1267
t (Observed value)	-0.6599
|t| (Critical value)	2.0926
DF	19
p-value (Two-tailed)	0.5172
alpha	0.05

Critical Value Method

The critical value method of testing a claim about two means is generally easier than the *P*-value method. With $n_1 = 12$ and $n_2 = 15$, the number of degrees of freedom is 11. In Table A-3 with df = 11 and $\alpha = 0.05$ in two tails, we get critical values of $t = \pm 2.201$. Technology provides $t = \pm 2.093$.

The test statistic of $t = -0.660$ falls between the critical values, so the test statistic is not in the critical region and we fail to reject the null hypothesis.



Example: Confidence Interval for Female and Male Course Evaluation Scores (1 of 5)

Using the previous data, construct a 95% confidence interval estimate of the difference between the mean course evaluation score for female professors and the mean course evaluation score for male professors.

Example: Confidence Interval for Female and Male Course Evaluation Scores (2 of 5)

Solution

Requirement check

Because we are using the same data from the previous example, the same requirement check applies here, so the requirements are satisfied.

Example: Confidence Interval for Female and Male Course Evaluation Scores (3 of 5)

Solution

We first find the value of the margin of error *E*. In Table A-3 with df = 11 and *α* = 0.05 in two tails, we get critical values of *t* = ± 2.201. (Technology can be used to find more accurate critical values of *t* = ± 2.093.

$$E = t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= 2.201 \sqrt{\frac{0.563001^2}{12} + \frac{0.395450^2}{15}} = 0.422452$$

Example: Confidence Interval for Female and Male Course Evaluation Scores

Solution

Using $E = 0.422452$, $\bar{x}_1 = 3.866667$, and $\bar{x}_2 = 3.993333$, we can now find the confidence interval as follows:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

$$-0.55 < (\mu_1 - \mu_2) < 0.30$$

If we use technology to obtain more accurate results, we get the confidence interval of $-0.53 < (\mu_1 - \mu_2) < 0.27$, so we can see that the confidence interval above is quite good.

Example: Confidence Interval for Female and Male Course Evaluation Scores (5 of 5)

Interpretation

We are 95% confident that the limits of −0.53 and 0.27 actually do contain the difference between the two population means. Because those limits contain 0, this confidence interval suggests that there is not a significant difference between the mean course evaluation score for female professors and the mean course evaluation score for male professors.

Alternate Methods: Assume that $\sigma_1 = \sigma_2$ and Pool the Sample Variances (1 of 2)

The **pooled estimate of σ^2** is denoted by s_p^2 and is a weighted average of s_1^2 and s_2^2, which is used in the test statistic for this case:

Test Statistic $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$$

where $s_p^2 = \dfrac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{(n_1 - 1) + (n_2 - 1)}$ (pooled sample variance)

Degrees of freedom : df = $n_1 + n_2 - 2$.

Alternate Methods: Assume that $\sigma_1 = \sigma_2$ and Pool the Sample Variances (2 of 2)

Confidence interval are found by evaluating

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

Margin of Error for Confidence Interval $E = t_{\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$

where $s_p{}^2 = \dfrac{(n_1 - 1)\, s_1{}^2 + (n_2 - 1)\, s_2{}^2}{(n_1 - 1) + (n_2 - 1)}$

Degrees of freedom : df = $n_1 + n_2 - 2$.

Alternate Methods: Assume When σ_1 and σ_2 Are Known (1 of 2)

In reality, the population standard deviations s_1 and s_2 are almost never known, but if they are somehow known, the test statistic and confidence interval are based on the normal distribution instead of the *t* distribution.

Test Statistic $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Pearson

Alternate Methods: Assume When σ_1 and σ_2 Are Known (2 of 2)

Confidence interval are found by evaluating

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

Margin of Error for Confidence Interval

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Recommended Strategy for Two Independent Means

Here is the recommended strategy for the methods of this section:

Assume that s_1 and s_2 are unknown, do not assume that $s_1 = s_2$, and use the test statistic and confidence interval given in Part 1 of this section.