Chapter 8

Hypothesis Testing with Two Samples

Chapter Outline

- 8.1 Testing the Difference Between Means (Large Independent Samples)
- 8.2 Testing the Difference Between Means (Small Independent Samples)
- 8.3 Testing the Difference Between Means (Dependent Samples)
- 8.4 Testing the Difference Between Proportions

Section 8.1

Testing the Difference Between Means (Large Independent Samples)

Section 8.1 Objectives

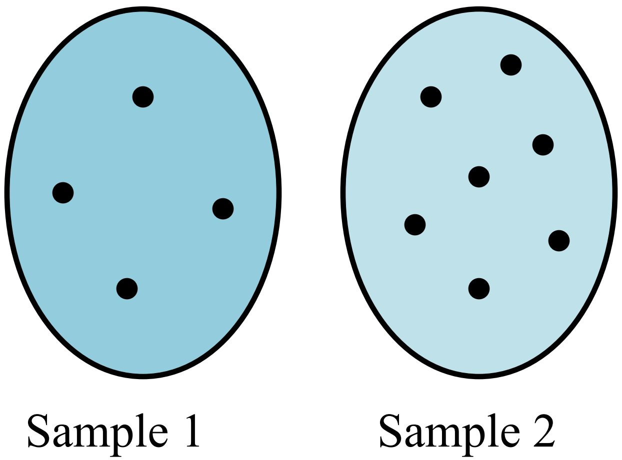
- Determine whether two samples are independent or dependent
- Perform a two-sample z-test for the difference between two means μ_1 and μ_2 using large independent samples

Two Sample Hypothesis Test

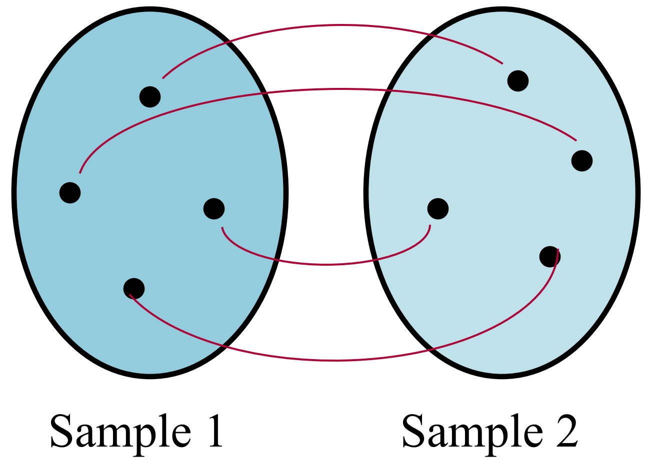
- Compares two parameters from two populations.
- Sampling methods:
  - **Independent Samples**
    - The sample selected from one population is not related to the sample selected from the second population.
  - **Dependent Samples** (paired or matched samples)
    - Each member of one sample corresponds to a member of the other sample.

Independent and Dependent Samples

Independent Samples



Dependent Samples



Example: Independent and Dependent Samples

Classify the pair of samples as independent or dependent.

- Sample 1: Resting heart rates of 35 individuals before drinking coffee.
- Sample 2: Resting heart rates of the same individuals after drinking two cups of coffee.

Solution:

Dependent Samples (The samples can be paired with respect to each individual)

Example: Independent and Dependent Samples

Classify the pair of samples as independent or dependent.

- Sample 1: Test scores for 35 statistics students.
- Sample 2: Test scores for 42 biology students who do not study statistics.

Solution:

Independent Samples (Not possible to form a pairing between the members of the samples; the sample sizes are different, and the data represent scores for different individuals.)

Two Sample Hypothesis Test with Independent Samples

1. **Null hypothesis H_0**
   - A statistical hypothesis that usually states there is no difference between the parameters of two populations.
   - Always contains the symbol =.
2. **Alternative hypothesis H_a**
   - A statistical hypothesis that is supported when H_0 is rejected.
   - Always contains the symbol $>$, $\neq$, or $<$.

Two Sample Hypothesis Test with Independent Samples

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 \neq \mu_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 > \mu_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 < \mu_2 \end{cases}$$

Regardless of which hypotheses you use, you always assume there is no difference between the population means, or $\mu_1 = \mu_2$.

Two Sample *z*-Test for the Difference Between Means

Three conditions are necessary to perform a *z*-test for the difference between two population means μ_1 and μ_2.

1. The samples must be randomly selected.
2. The samples must be independent.
3. Each sample size must be at least 30, or, if not, each population must have a normal distribution with a known standard deviation.

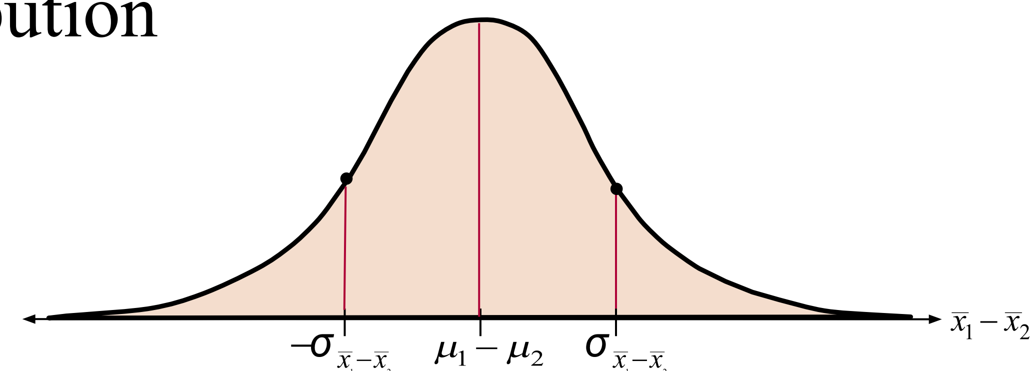
Two Sample *z*-Test for the Difference Between Means

If these requirements are met, the sampling distribution for $\bar{x}_1 - \bar{x}_2$ (the difference of the sample means) is a normal distribution with

Mean: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$

Standard error: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Sampling distribution for $\bar{x}_1 - \bar{x}_2$:



Two Sample *z*-Test for the Difference Between Means

- **Test statistic** is $\bar{x}_1 - \bar{x}_2$
- The **standardized test statistic** is

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

- When the samples are large, you can use s_1 and s_2 in place of σ_1 and σ_2. If the samples are not large, you can still use a two-sample z-test, provided the populations are normally distributed and the population standard deviations are known.

Using a Two-Sample z-Test for the Difference Between Means (Large Independent Samples)

In Words	*In Symbols*
1. State the claim mathematically. Identify the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Sketch the sampling distribution.	
4. Determine the critical value(s).	Use Table 4 in Appendix B.
5. Determine the rejection region(s).	

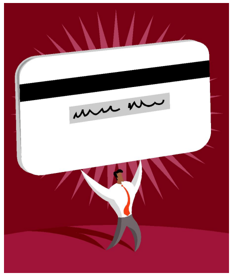
Using a Two-Sample z-Test for the Difference Between Means (Large Independent Samples)

In Words	*In Symbols*
6. Find the standardized test statistic.	$z = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{x}_1 - \overline{x}_2}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If z is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

Example: Two-Sample *z*-Test for the Difference Between Means

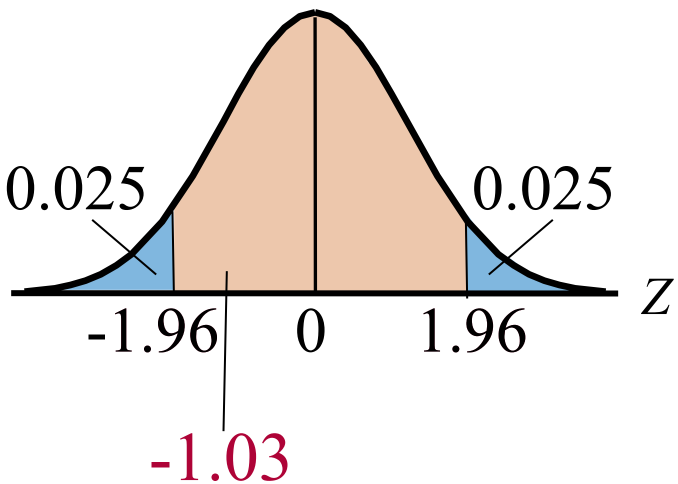
A consumer education organization claims that there is a difference in the mean credit card debt of males and females in the United States. The results of a random survey of 200 individuals from each group are shown below. The two samples are independent. Do the results support the organization's claim? Use α = 0.05.

Females (1)	Males (2)
$\bar{x}_1 = \$2290$	$\bar{x}_2 = \$2370$
$s_1 = \$750$	$s_2 = \$800$
$n_1 = 200$	$n_2 = 200$



Solution: Two-Sample *z*-Test for the Difference Between Means

- H_0: $\mu_1 = \mu_2$
- H_a: $\mu_1 \neq \mu_2$
- $\alpha = 0.05$
- $n_1 = 200$, $n_2 = 200$
- **Rejection Region:**



- **Test Statistic:**

$$z = \frac{(2290 - 2370) - 0}{\sqrt{\frac{750^2}{200} + \frac{800^2}{200}}} = -1.03$$

- **Decision: Fail to Reject H_0**

At the 5% level of significance, there is not enough evidence to support the organization's claim that there is a difference in the mean credit card debt of males and females.

Example: Using Technology to Perform a Two-Sample *z*-Test

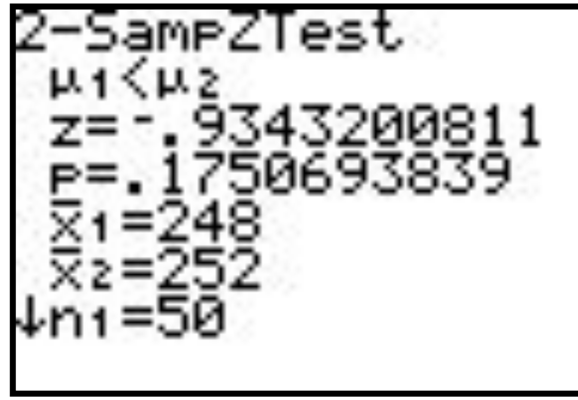
The American Automobile Association claims that the average daily cost for meals and lodging for vacationing in Texas is less than the same average costs for vacationing in Virginia. The table shows the results of a random survey of vacationers in each state. The two samples are independent. At $\alpha = 0.01$, is there enough evidence to support the claim?

Texas (1)	Virginia (2)
$\bar{x}_1 = \$248$	$\bar{x}_2 = \$252$
$s_1 = \$15$	$s_2 = \$22$
$n_1 = 50$	$n_2 = 35$

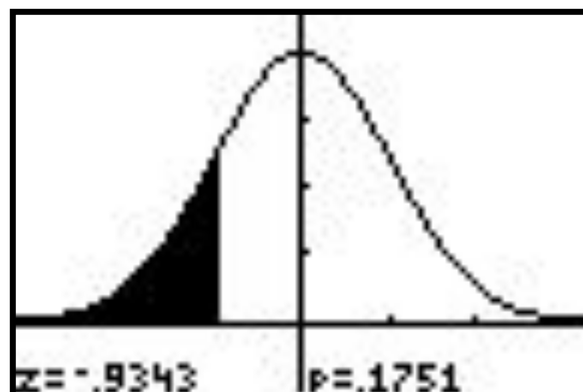
Solution: Using Technology to Perform a Two-Sample *z*-Test

- H_0: $\mu_1 = \mu_2$
- H_a: $\mu_1 < \mu_2$

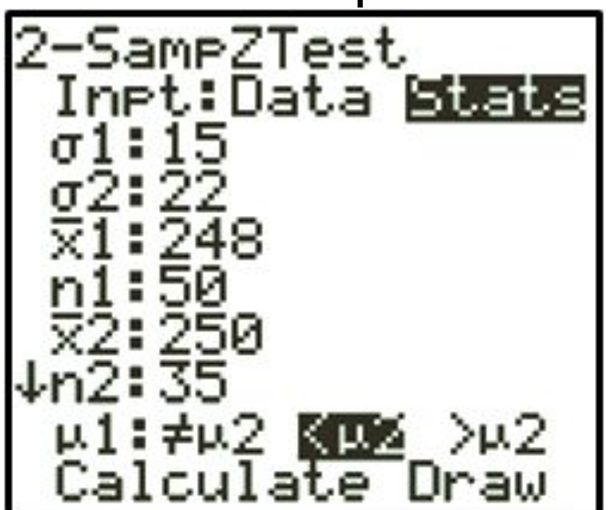
Calculate:



Draw:

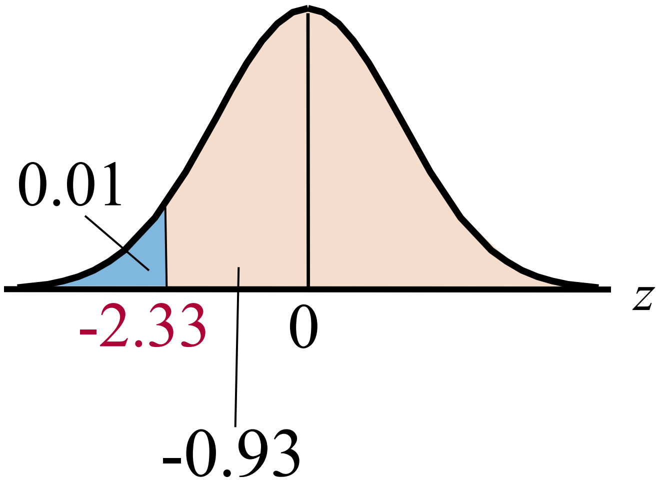


TI-83/84set up:



Solution: Using Technology to Perform a Two-Sample *z*-Test

- **Rejection Region:**



- **Decision:** **Fail to Reject H_0**
  At the 1% level of significance, there is not enough evidence to support the American Automobile Association's claim.

Section 8.1 Summary

- Determined whether two samples are independent or dependent
- Performed a two-sample *z*-test for the difference between two means μ_1 and μ_2 using large independent samples

Section 8.2

Testing the Difference Between Means (Small Independent Samples)

Section 8.2 Objectives

- Perform a *t*-test for the difference between two means μ_1 and μ_2 using small independent samples

Two Sample *t*-Test for the Difference Between Means

- If samples of size less than 30 are taken from normally-distributed populations, a *t*-test may be used to test the difference between the population means μ_1 and μ_2.
- Three conditions are necessary to use a *t*-test for small independent samples.

1. The samples must be randomly selected.
2. The samples must be independent.
3. Each population must have a normal distribution.

Two Sample *t*-Test for the Difference Between Means

- The **standardized test statistic** is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

- The standard error and the degrees of freedom of the sampling distribution depend on whether the population variances σ_1^2 and σ_2^2 are equal.

Two Sample *t*-Test for the Difference Between Means

- **Variances are equal**
  - Information from the two samples is combined to calculate a **pooled estimate of the standard deviation** $\hat{\sigma}$**.**

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

  - The standard error for the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

  - d.f.$= n_1 + n_2 - 2$

Two Sample *t*-Test for the Difference Between Means

- **Variances are not equal**
  - If the population variances are not equal, then the standard error is

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

  - d.f = smaller of $n_1 - 1$ or $n_2 - 1$

Normal or *t*-Distribution?

Are both sample sizes at least 30?

Yes → Use the z-test.

No ↓

Are both populations normally distributed?

No → You cannot use the z-test or the t-test.

Yes ↓

Are both population standard deviations known?

Yes ↓ Use the z-test.

No → Are the population variances equal?

Yes → Use the t-test with

$$\sigma_{\bar{x}_1-\bar{x}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\text{d.f} = n_1 + n_2 - 2.$$

No ↓ Use the t-test with

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$\text{d.f} = \text{smaller of } n_1 - 1 \text{ or } n_2 - 1.$$

Two-Sample *t*-Test for the Difference Between Means (Small Independent Samples)

In Words	*In Symbols*
1. State the claim mathematically. Identify the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom and sketch the sampling distribution.	d.f. = $n_1 + n_2 - 2$ or d.f. = smaller of $n_1 - 1$ or $n_2 - 1$.
4. Determine the critical value(s).	Use Table 5 in Appendix B.

Two-Sample *t*-Test for the Difference Between Means (Small Independent Samples)

In Words	*In Symbols*
5. Determine the rejection region(s).	
6. Find the standardized test statistic.	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

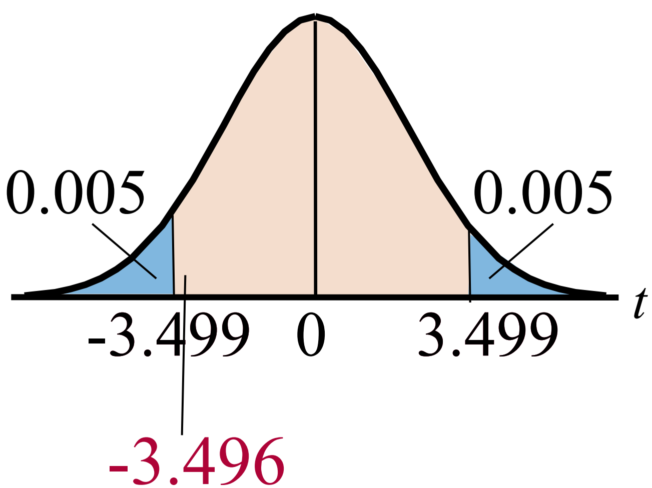
Example: Two-Sample *t*-Test for the Difference Between Means

The braking distances of 8 Volkswagen GTIs and 10 Ford Focuses were tested when traveling at 60 miles per hour on dry pavement. The results are shown below. Can you conclude that there is a difference in the mean braking distances of the two types of cars? Use α = 0.01. Assume the populations are normally distributed and the population variances are not equal. *(Adapted from Consumer Reports)*

GTI (1)	Focus (2)
$\bar{x}_1 = 134$ ft	$\bar{x}_2 = 143$ ft
$s_1 = 6.9$ ft	$s_2 = 2.6$ ft
$n_1 = 8$	$n_2 = 10$

Solution: Two-Sample *t*-Test for the Difference Between Means

- H_0: $\mu_1 = \mu_2$
- H_a: $\mu_1 \neq \mu_2$
- $\alpha =$ **0.01**
- **d.f. = 8 – 1 = 7**
- **Rejection Region:**



- **Test Statistic:**

$$t = \frac{(134 - 143) - 0}{\sqrt{\frac{6.9^2}{8} + \frac{2.6^2}{10}}} = -3.496$$

- **Decision:** **Fail to Reject H_0**

At the 1% level of significance, there is not enough evidence to conclude that the mean braking distances of the cars are different.

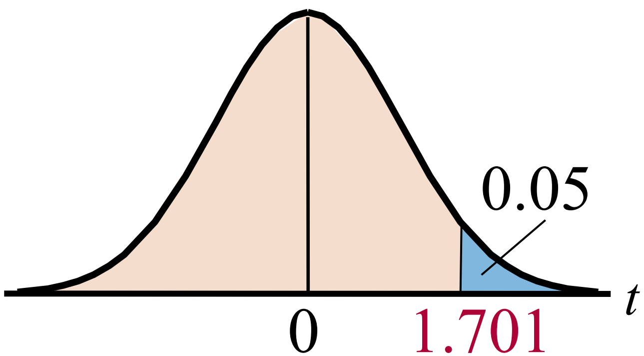
Example: Two-Sample *t*-Test for the Difference Between Means

A manufacturer claims that the calling range (in feet) of its 2.4-GHz cordless telephone is greater than that of its leading competitor. You perform a study using 14 randomly selected phones from the manufacturer and 16 randomly selected similar phones from its competitor. The results are shown below. At $\alpha = 0.05$, can you support the manufacturer's claim? Assume the populations are normally distributed and the population variances are equal.

Manufacturer (1)	Competition (2)
$\bar{x}_1 = 1275$ ft $s_1 = 45$ ft $n_1 = 14$	$\bar{x}_2 = 1250$ ft $s_2 = 30$ ft $n_2 = 16$

Solution: Two-Sample *t*-Test for the Difference Between Means

- H_0: $\mu_1 = \mu_2$
- H_a: $\mu_1 > \mu_2$
- $\alpha = 0.05$
- d.f. = $14 + 16 - 2 = 28$
- Rejection Region:



- Test Statistic:
- Decision:

Solution: Two-Sample *t*-Test for the Difference Between Means

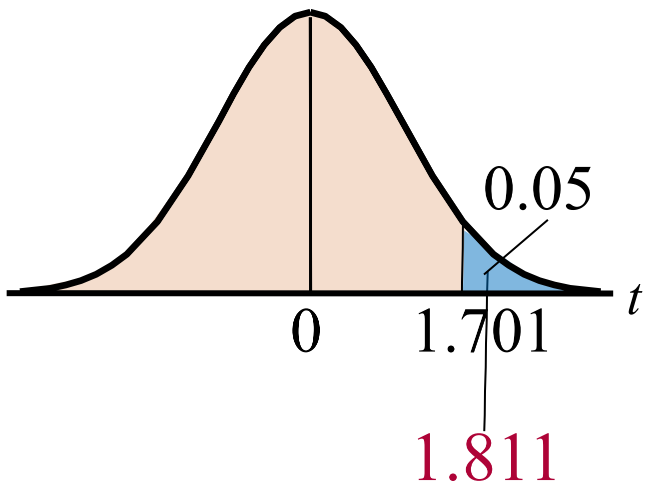
$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= \sqrt{\frac{(14-1)(45)^2+(16-1)(30)^2}{14+16-2}}\cdot\sqrt{\frac{1}{14}+\frac{1}{16}} \approx 13.8018$$

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} = \frac{(1275-1250)-0}{13.8018} \approx 1.811$$

Solution: Two-Sample *t*-Test for the Difference Between Means

- H_0: $\mu_1 = \mu_2$
- H_a: $\mu_1 > \mu_2$
- $\alpha =$ **0.05**
- **d.f. = 14 + 16 – 2 = 28**
- **Rejection Region:**



- **Test Statistic:**

$$t = 1.811$$

- **Decision: Reject H_0**

At the 5% level of significance, there is enough evidence to support the manufacturer's claim that its phone has a greater calling range than its competitors.

Section 8.2 Summary

- Performed a *t*-test for the difference between two means μ_1 and μ_2 using small independent samples

Section 8.3

Testing the Difference Between Means (Dependent Samples)

Section 8.3 Objectives

- Perform a *t*-test to test the mean of the difference for a population of paired data

t-Test for the Difference Between Means

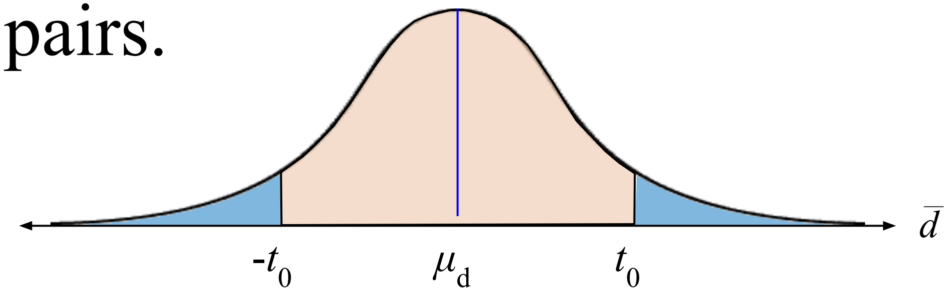
- To perform a two-sample hypothesis test with dependent samples, the difference between each data pair is first found:
  - $d = x_1 - x_2$ Difference between entries for a data pair
- The test statistic is the mean $\bar{d}$ of these differences.
  - $\bar{d} = \dfrac{\sum d}{n}$ Mean of the differences between paired data entries in the dependent samples

t-Test for the Difference Between Means

Three conditions are required to conduct the test.

1. The samples must be randomly selected.
2. The samples must be dependent (paired).
3. Both populations must be normally distributed.

If these conditions are met, then the sampling distribution for $\bar{d}$ is approximated by a *t*-distribution with $n - 1$ degrees of freedom, where n is the number of data pairs.



Symbols used for the *t*-Test for μ_d

Symbol	Description
n	The number of pairs of data
d	The difference between entries for a data pair, $d = x_1 - x_2$
μ_d	The hypothesized mean of the differences of paired data in the population

Symbols used for the *t*-Test for μ_d

Symbol	Description
$\bar{d}$	The mean of the differences between the paired data entries in the dependent samples $\bar{d} = \dfrac{\sum d}{n}$
s_d	The standard deviation of the differences between the paired data entries in the dependent samples $s_d = \sqrt{\dfrac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\dfrac{\Sigma d^2 - \dfrac{(\Sigma d)^2}{n}}{n-1}}$

t-Test for the Difference Between Means

- The **test statistic** is

$$\bar{d} = \frac{\sum d}{n}$$

- The **standardized test statistic** is

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

- The degrees of freedom are

$$\text{d.f.} = n - 1$$

t-Test for the Difference Between Means (Dependent Samples)

In Words	*In Symbols*
1. State the claim mathematically. Identify the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom and sketch the sampling distribution.	d.f. $= n - 1$
4. Determine the critical value(s).	Use Table 5 in Appendix B if $n > 29$ use the last row (∞) .

t-Test for the Difference Between Means (Dependent Samples)

In Words	*In Symbols*
5. Determine the rejection region(s).	
6. Calculate $\bar{d}$ and s_d. Use a table.	$\bar{d} = \dfrac{\sum d}{n}$ $s_d = \sqrt{\dfrac{\sum(d-\bar{d})^2}{n-1}} = \sqrt{\dfrac{\Sigma d^2 - \dfrac{(\Sigma d)^2}{n}}{n-1}}$
7. Find the standardized test statistic.	$t = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$

t-Test for the Difference Between Means (Dependent Samples)

In Words	*In Symbols*
8. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
9. Interpret the decision in the context of the original claim.	

Example: *t*-Test for the Difference Between Means

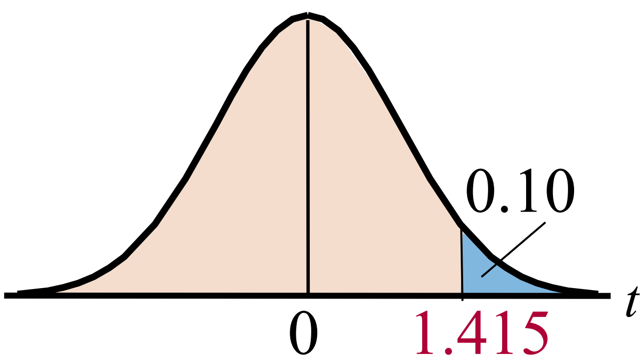
A golf club manufacturer claims that golfers can lower their scores by using the manufacturer's newly designed golf clubs. Eight golfers are randomly selected, and each is asked to give his or her most recent score. After using the new clubs for one month, the golfers are again asked to give their most recent score. The scores for each golfer are shown in the table. Assuming the golf scores are normally distributed, is there enough evidence to support the manufacturer's claim at $\alpha = 0.10$?

Golfer	1	2	3	4	5	6	7	8
Score (old)	89	84	96	82	74	92	85	91
Score (new)	83	83	92	84	76	91	80	91

Solution: Two-Sample *t*-Test for the Difference Between Means

d = (old score) – (new score)

- **H_0: $\mu_d \le 0$**
- **H_a: $\mu_d > 0$**
- **α = 0.10**
- **d.f. = 8 – 1 = 7**
- **Rejection Region:**



- **Test Statistic:**
- **Decision:**

Solution: Two-Sample *t*-Test for the Difference Between Means

d = (old score) – (new score)

Old	New	d	d^2
89	83	6	36
84	83	1	1
96	92	4	16
82	84	–2	4
74	76	–2	4
92	91	1	1
85	80	5	25
91	91	0	0
		$\Sigma = 13$	$\Sigma = 87$

$$\bar{d} = \frac{\sum d}{n} = \frac{13}{8} = 1.625$$

$$s_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}}$$

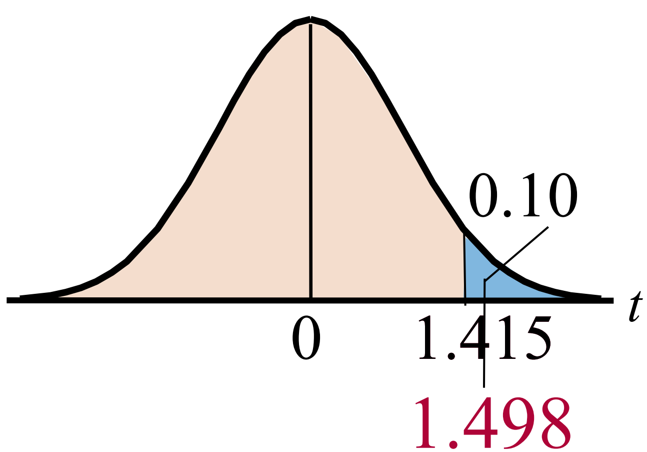
$$= \sqrt{\frac{87 - \frac{(13)^2}{8}}{8-1}}$$

$$\approx 3.0677$$

Solution: Two-Sample *t*-Test for the Difference Between Means

d = (old score) – (new score)

- H_0: $\mu_d \le 0$
- H_a: $\mu_d > 0$
- $\alpha = 0.10$
- d.f. = $8 - 1 = 7$
- **Rejection Region:**



- **Test Statistic:**

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \approx \frac{1.625 - 0}{3.0677 / \sqrt{8}} \approx 1.498$$

- **Decision:** Reject H_0

At the 10% level of significance, the results of this test indicate that after the golfers used the new clubs, their scores were significantly lower.

Section 8.3 Summary

- Performed a *t*-test to test the mean of the difference for a population of paired data

Section 8.4

Testing the Difference Between Proportions

Section 8.4 Objectives

- Perform a z-test for the difference between two population proportions p_1 and p_2

Two-Sample *z*-Test for Proportions

- Used to test the difference between two population proportions, p_1 and p_2.
- Three conditions are required to conduct the test.
  1. The samples must be randomly selected.
  2. The samples must be independent.
  3. The samples must be large enough to use a normal sampling distribution. That is, $n_1p_1 \geq 5$, $n_1q_1 \geq 5$, $n_2p_2 \geq 5$, and $n_2q_2 \geq 5$.

Two-Sample *z*-Test for the Difference Between Proportions

- If these conditions are met, then the sampling distribution for $\hat{p}_1 - \hat{p}_2$ is a normal distribution
- Mean: $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$
- A weighted estimate of p_1 and p_2 can be found by using

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}, \text{ where } x_1 = n_1\hat{p}_1 \text{ and } x_2 = n_2\hat{p}_2$$

- Standard error:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Two-Sample *z*-Test for the Difference Between Proportions

- The **test statistic** is $\hat{p}_1 - \hat{p}_2$
- The **standardized test statistic** is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ and } \bar{q} = 1 - \bar{p}$$

Note: $n_1\bar{p},\ n_1\bar{q},\ n_2\bar{p},$ and $n_2\bar{q}$ must be at least 5.

Two-Sample *z*-Test for the Difference Between Proportions

In Words	*In Symbols*
1. State the claim. Identify the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Determine the critical value(s).	Use Table 4 in Appendix B.
4. Determine the rejection region(s).	
5. Find the weighted estimate of p_1 and p_2.	$\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

Two-Sample z-Test for the Difference Between Proportions

In Words	*In Symbols*
6. Find the standardized test statistic.	$z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If z is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

Example: Two-Sample *z*-Test for the Difference Between Proportions

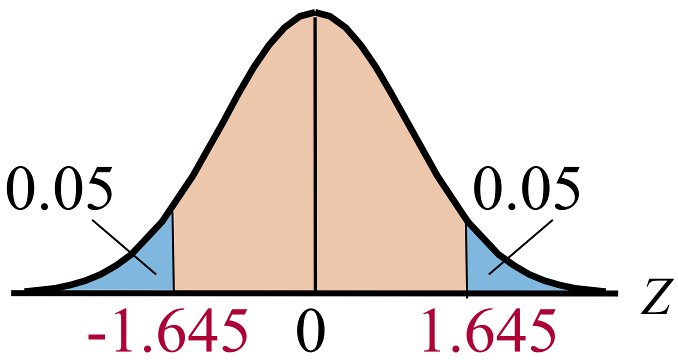
In a study of 200 randomly selected adult female and 250 randomly selected adult male Internet users, 30% of the females and 38% of the males said that they plan to shop online at least once during the next month. At $\alpha = 0.10$ test the claim that there is a difference between the proportion of female and the proportion of male Internet users who plan to shop online.

Solution:

1 = Females   2 = Males

Solution: Two-Sample *z*-Test for the Difference Between Means

- H_0: $p_1 = p_2$
- H_a: $p_1 \neq p_2$
- $\alpha = 0.10$
- $n_1 = 200$, $n_2 = 250$
- Rejection Region:



- Test Statistic:
- Decision:

Solution: Two-Sample *z*-Test for the Difference Between Means

$$x_1 = n_1 \hat{p}_1 = 60 \qquad x_2 = n_2 \hat{p}_2 = 95$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{60 + 95}{200 + 250} \approx 0.3444$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.3444 = 0.6556$$

Note:

$$n_1\bar{p} = 200(0.3444) \geq 5 \qquad n_1\bar{q} = 200(0.6556) \geq 5$$

$$n_2\bar{p} = 250(0.3444) \geq 5 \qquad n_2\bar{q} = 250(0.6556) \geq 5$$

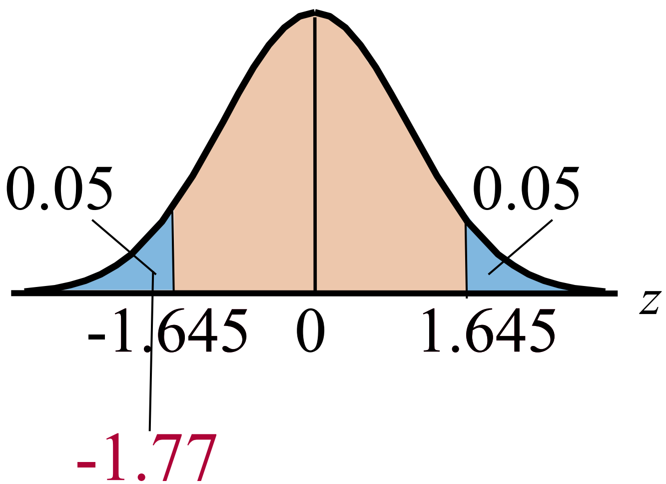
Solution: Two-Sample *z*-Test for the Difference Between Means

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\overline{q} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0.30 - 0.38) - (0)}{\sqrt{(0.3444) \cdot (0.6556) \cdot \left(\frac{1}{200} + \frac{1}{250}\right)}}$$

$$\approx -1.77$$

Solution: Two-Sample *z*-Test for the Difference Between Means

- H_0: $p_1 = p_2$
- H_a: $p_1 \neq p_2$
- $\alpha = 0.10$
- $n_1 = 200$, $n_2 = 250$
- **Rejection Region:**



- **Test Statistic:**

$$z = -1.77$$

- **Decision:** Reject H_0

At the 10% level of significance, there is enough evidence to conclude that there is a difference between the proportion of female and the proportion of male Internet users who plan to shop online.

Example: Two-Sample *z*-Test for the Difference Between Proportions

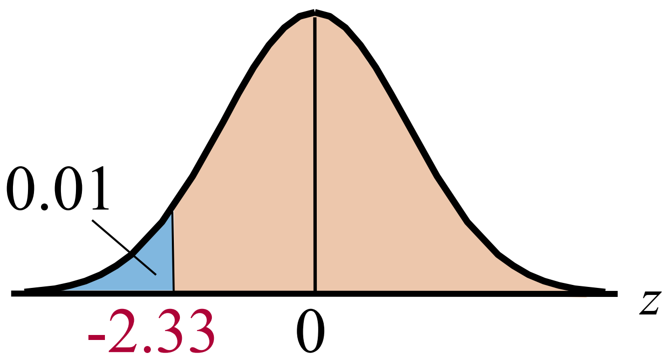
A medical research team conducted a study to test the effect of a cholesterol reducing medication. At the end of the study, the researchers found that of the 4700 randomly selected subjects who took the medication, 301 died of heart disease. Of the 4300 randomly selected subjects who took a placebo, 357 died of heart disease. At $\alpha = 0.01$ can you conclude that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo? *(Adapted from New England Journal of Medicine)*

Solution:

1 = Medication 2 = Placebo

Solution: Two-Sample *z*-Test for the Difference Between Means

- H_0: $p_1 \geq p_2$
- H_a: $p_1 < p_2$
- $\alpha = 0.01$
- $n_1 = 4700$, $n_2 = 4300$
- **Rejection Region:**



- **Test Statistic:**
- **Decision:**

Solution: Two-Sample *z*-Test for the Difference Between Means

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{301}{4700} = 0.064 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{357}{4300} = 0.083$$

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{301 + 357}{4700 + 4300} \approx 0.0731$$

$$\overline{q} = 1 - \overline{p} = 1 - 0.0731 = 0.9269$$

Note:

$$n_1\overline{p} = 4700(0.0731) \geq 5 \qquad n_1\overline{q} = 4700(0.9269) \geq 5$$

$$n_2\overline{p} = 4300(0.0731) \geq 5 \qquad n_2\overline{q} = 4300(0.9269) \geq 5$$

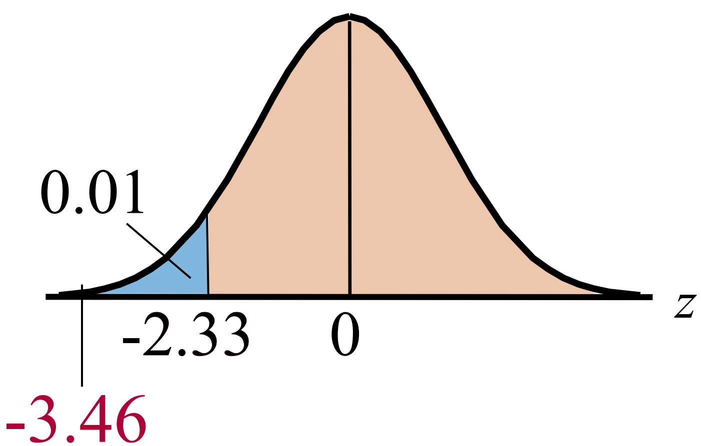
Solution: Two-Sample *z*-Test for the Difference Between Means

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0.064 - 0.083) - (0)}{\sqrt{(0.0731)\cdot(0.9269)\cdot\left(\frac{1}{4700} + \frac{1}{4300}\right)}}$$

$$\approx -3.46$$

Solution: Two-Sample *z*-Test for the Difference Between Means

- H_0: $p_1 \geq p_2$
- H_a: $p_1 < p_2$
- $\alpha = 0.01$
- $n_1 = 4700$, $n_2 = 4300$
- **Rejection Region:**



- **Test Statistic:**

$$z = -3.46$$

- **Decision: Reject H_0**

At the 1% level of significance, there is enough evidence to conclude that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo.

Section 8.4 Summary

- Performed a z-test for the difference between two population proportions p_1 and p_2