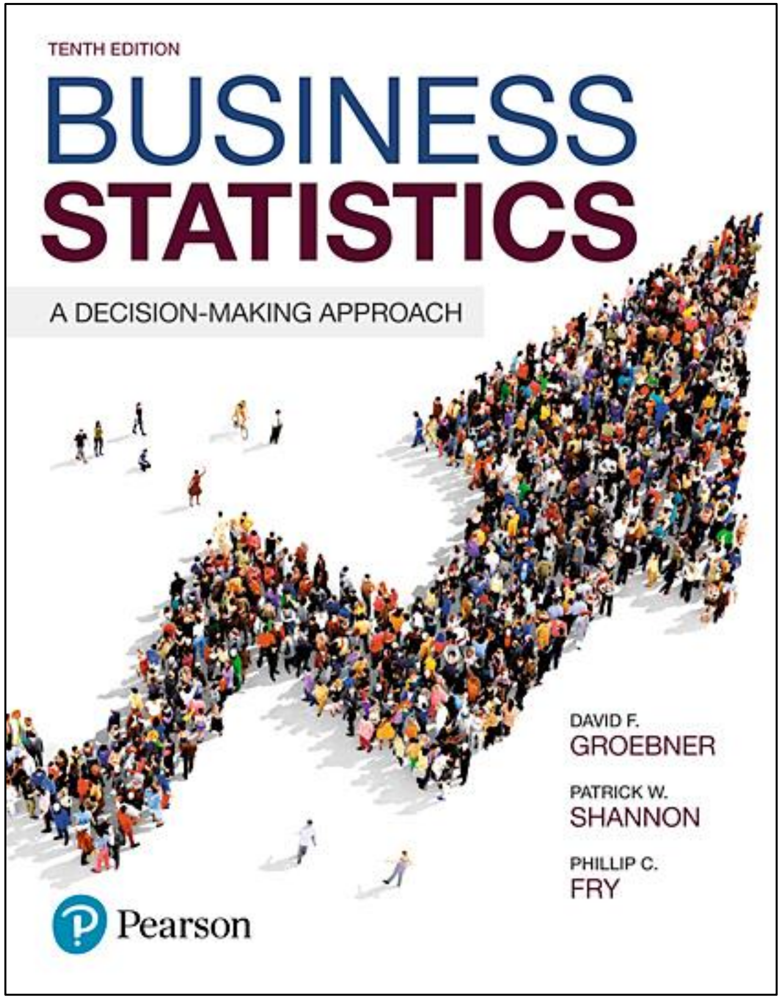
Business Statistics: A Decision-Making Approach

Tenth Edition



Chapter 10

Estimating and Hypothesis Testing

Objectives

1. Discuss the logic behind and demonstrate the techniques for using independent samples to test hypotheses and develop interval estimates for the difference between two popular means.
2. Develop confidence interval estimates and conduct hypothesis tests for the difference between two population means for paired samples
3. Carry out hypothesis tests and establish interval estimates, using sample data, for the difference between two population proportions.

10.1 Estimation the Difference Between Two Population Means Using Independent Samples

- Independent Samples
  - Samples selected from two or more populations in such a way that the occurrence of values in one sample has no influence on the probability of the occurrence of values in the other sample(s).

Estimation for Two Population Means

The population standard deviations are known

The population standard deviations are unknown

Estimating the Difference Mu Sub 1 Minus Mu Sub 2 Sigma Sub 1 and Sigma Sub 2 Are Known

- The best point estimate for the difference of two population means $\mu_1 - \mu_2$ is:

$$\text{Point estimate} = \bar{x}_1 - \bar{x}_2$$

- Standard Error of $\bar{x}_1 - \bar{x}_2$:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

σ_1^2 - Variance of population 1

σ_2^2 - Variance of population 2

n_1 and n_2 - Sample sizes of populations 1 and 2

Confidence Interval Estimate for Mu Sub 1 Minus Mu Sub 2 with Sigma Sub 1 and Sigma Sub 2 Known

General Format

$$\text{Point Estimate} \pm \left(\text{Critical Value}\right)(\text{Standard Error})$$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n_1 + n_2}}$$

- z-values for several of the most used confidence levels:

Confidence Level	Critical z-value
80%	z = 1.28
90%	z = 1.645
95%	z = 1.96
99%	z = 2.575

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Known

- Step 1: Define the population parameter of interest and select independent samples from the two populations
- Step 2: Specify the desired confidence level
- Step 3: Compute the point estimate
- Step 4: Determine the standard error of the sampling distribution
- Step 5: Determine the critical value, z, from the standard normal table
- Step 6: Develop the confidence interval estimate

Confidence Interval Estimate for Mu Sub 1 Minus Mu Sub 2 with Sigma Sub 1 and Sigma Sub 2 Known - Example (1 of 2)

The broker for a real estate agency wishes to estimate the difference in mean number of days a home tales to sell for one-story versus two-story. A 95% confidence level is to be used with sample sizes for one and two-story homes equal to $n_1 = 80$ and $n_2 = 100$ respectively. The standard deviation for one-story homes is 10 days and the standard deviation for two-story homes in 14 days. The sample mean for one-story is 96 days and for two-story is 75 days.

$$\text{Point Estimate} = (\bar{x}_1 - \bar{x}_2) = (96 - 75) = 21$$

$$\text{St. Error} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n_1 + n_2}} = \sqrt{\frac{10^2 + 14^2}{80 + 100}} = 1.2824$$

Confidence Interval Estimate for Mu Sub 1 Minus Mu Sub 2 with Sigma Sub 1 and Sigma Sub 2 Known - Example (2 of 2)

General Format

$$\text{Point Estimate} \pm \left(\text{Critical Value}\right)(\text{Standard Error})$$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n_1 + n_2}}$$

$$21 \pm (1.96)(1.2824)$$

$$21 \pm 2.3175$$

Margin of Error

$$18.6825 \text{ --------- } 23.3175$$

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Unknown (1 of 3)

- Assumptions:
  - The populations are normally distributed
  - The populations have equal variances
  - The samples are independent
- Confidence interval estimate should be developed using the *t*-distribution

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Unknown (2 of 3)

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s_p = \sqrt{\frac{\left(n_1 - 1\right) s_1^{\,2} + \left(n_2 - 1\right) s_2^{\,2}}{n_1 + n_2 - 2}}$$ - Pooled standard deviation

t - Critical *t*-value from the *t*-distribution, with degrees of freedom equal to $n_1 + n_2 - 2$

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Unknown (3 of 3)

- Step 1: Define the population parameter of interest and select independent samples from the two populations
- Step 2: Specify the confidence level
- Step 3: Compute the point estimate
- Step 4: Determine the standard error of sampling distribution
- Step 5: Determine the critical value, *t*, from the *t*-distribution table
- Step 6: Develop a confidence interval

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Unknown - Example (1 of 2)

An insurance company wishes to estimate the difference in mean damage to cars that crash into a barricade at 20 mph with a new bumper system versus the older bumper system. A random sample on 8 cars with the new bumper system provided a mean damage equal to $3,950 and a sample standard deviation equal to $600. A random sample of 10 cars with the older bumper system provided a mean damage of $3,475 and a sample standard deviation equal to $650. Develop a 95% confidence interval estimate.

$$\text{Point Estimate} = (\bar{x}_1 - \bar{x}_2) = (\$3,950 - \$3,475) = \$475$$

$$\text{St. Error} = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}};\ \text{where } s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_1-1)s_1^2}{n_1+n_2-2}} = \sqrt{\frac{(8-1)600^2+(10-1)650^2}{8+10-2}} = \$628.61$$

$$\text{St. Error} = 629.61\sqrt{\frac{1}{8}+\frac{1}{10}} = \$31.87$$

Confidence Interval for Mu Sub 1 Minus Mu Sub 2, Sigma Sub 1 and Sigma Sub 2 Are Unknown - Example (2 of 2)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.95, df=16} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\$475 \pm (2.1199)(\$31.87)$$

$$\$475 \pm 67.56$$

$407.44 ----------- $542.56

Based on the sample data with 95% confidence we conclude that the difference between mean damage for the two bumpers is between $407.44 and $542.56.

Sigma Sub 1 and Sigma Sub 2 Are Unknown and Not Equal (1 of 2)

- Confidence Interval for $\mu_1 - \mu_2$:

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

t - Critical *t*-value from the *t*-distribution, with degrees of freedom equal to:

$$df = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\left(\frac{\left(\frac{{s_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{{s_2}^2}{n_2}\right)^2}{n_2 - 1}\right)}$$

Sigma Sub 1 and Sigma Sub 2 Are Unknown and Not Equal (2 of 2)

- Step 1: Define the population parameter of interest and select independent samples from the two populations
- Step 2: Specify the desired confidence level
- Step 3: Compute the point estimate
- Step 4: Determine the standard error of the sampling distribution
- Step 5: Calculate the degrees of freedom and determine the critical value, *t*, from the *t*-distribution table
- Step 6: Develop the confidence interval estimate

Estimating Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Unknown and Not Equal - Example (1 of 3)

A major bank is planning a new marketing campaign in which low interest rates are being used to entice people to spend more money with the bank's charge card. The bank is interested in estimating the difference between the mean spending by married versus unmarried card holders. A random sample of 30 unmarried and 25 married customers was selected. They wish to develop a 95% confidence interval estimate for the difference in populations means but do not believe that the population standard deviations are equal. The following sample results were observed:

	Unmarried	**Married**
Mean	\$455.10	\$268.90
St. Dev.	\$102.40	\$77.25

Estimating Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Unknown and Not Equal - Example (2 of 3)

The confidence interval estimate is computed using

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}$$

The standard error is calculated as

$$\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}} = \sqrt{\frac{102.40^2}{30} + \frac{77.25^2}{25}} = 24.25$$

$$df = \frac{\left(\frac{s_1{}^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)} = \frac{\left(\frac{102.40^2}{30} + \frac{77.25^2}{25}\right)^2}{\left(\frac{\left(\frac{102.40^2}{30}\right)^2}{29} + \frac{\left(\frac{77.25^2}{25}\right)^2}{24}\right)} = \frac{346{,}011.98}{6{,}586.81} = 52.53$$

Estimating Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Unknown and Not Equal - Example (3 of 3)

The confidence interval estimate is computed using

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

The Excel 2016 function for determining the exact critical *t*-value is

= T.INV.2T(1 – 0.95,52)

Then the interval estimate is

$$(\$455.10 - \$268.90) \pm 2.0086\sqrt{\frac{102.40^2}{30} + \frac{77.25^2}{25}}$$

$$\$186.20 \pm \$48.72$$

$$\$137.48 \le (\mu_1 - \mu_2) \le \$234.92$$

$$\$137.48 \underline{\hspace{4cm}} \$234.92$$

10.2 Hypothesis Tests for Two Population Means Using Independent Samples

- Some situations require to test whether two populations have equal means or whether one population mean is larger (or smaller) then another
- These are hypothesis-testing applications

Hypothesis Tests for Two Population Means

The population standard deviations are known and the samples are independent

The population standard deviations are unknown and the samples are independent

The Hypothesis-Testing Process for Two Population Means (1 of 2)

- Step 1: Specify the population parameter of interest
- Step 2: Formulate the appropriate null and alternative hypotheses. The null hypothesis should contain the equality

Variations in Hypothesis Testing

Two-tailed Test	One-tailed Test (lower)	One-tailed Test (upper)
$H_0: \mu_1 - \mu_2 = 0$	$H_0: \mu_1 - \mu_2 \geq 0$	$H_0: \mu_1 - \mu_2 \leq 0$
$H_A: \mu_1 - \mu_2 \neq 0$	$H_A: \mu_1 - \mu_2 < 0$	$H_A: \mu_1 - \mu_2 > 0$
or	or	or
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 \geq \mu_2$	$H_0: \mu_1 \leq \mu_2$
$H_A: \mu_1 \neq \mu_2$	$H_A: \mu_1 < \mu_2$	$H_A: \mu_1 > \mu_2$

The Hypothesis-Testing Process for Two Population Means (2 of 2)

- Step 3: Specify the significance level (α) for testing the hypothesis. α is the maximum allowable probability of committing a Type I statistical error
- Step 4: Determine the rejection region and develop the decision rule
- Step 5: Compute the test statistic or the *p*-value. The simple random samples from each population should be selected and the sample means need to be computed
- Step 6: Reach a decision. Apply the decision rule to determine whether to reject the null hypothesis.
- Step 7: Draw a conclusion

Hypothesis-Testing for Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Known-Example (1 of 2)

A company has designed two new highly automated methods for assembling a cell phone at a Chinese plant in Shanghai. One measure of importance is the mean time it takes to insert the battery into the battery dock in the phone. Engineers believe that the mean time for method 1 is less than the mean time for method 2. To test whether this is correct, 100 phones are assembled using each of the methods and the battery install time is recorded. $\sigma_1 = 0.025$ seconds and $\sigma_2 = 0.034$ seconds. The sample data give the following:

$\bar{x}_1 = 0.501$ sec.

$\bar{x}_2 = 0.509$ sec.

$$H_o: \mu_1 \geq \mu_2 \qquad H_o: \mu_1 - \mu_2 \geq 0$$

$$H_A: \mu_1 < \mu_2 \quad \text{or} \quad H_A: \mu_1 - \mu_2 < 0$$

$$\alpha = 0.05$$

One-tailed, lower tail test

Hypothesis-Testing for Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Known-Example (2 of 2)

$$\bar{x}_1 = 0.501 \text{ sec.}$$

$$\bar{x}_2 = 0.509 \text{ sec.}$$

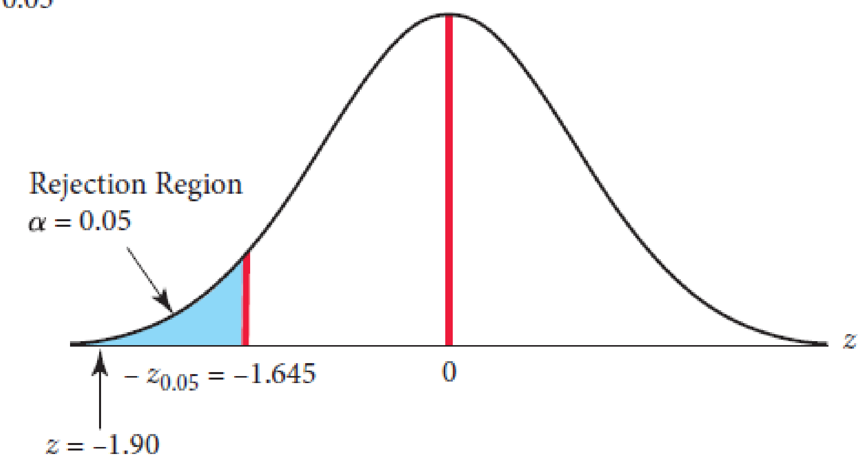
$$\sigma_1 = 0.025 \text{ sec.}$$

$$\sigma_2 = 0.034 \text{ sec.}$$

$H_0: \mu_1 \geq \mu_2$ or $H_0: \mu_1 - \mu_2 \geq 0$

$H_A: \mu_1 < \mu_2$ or $H_A: \mu_1 - \mu_2 < 0$

$\alpha = 0.05$



Test Statistic:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(0.501 - 0.509) - 0}{\sqrt{\frac{0.025^2}{100} + \frac{0.034^2}{100}}} = -1.90$$

Decision Rule:

Since $z = -1.90 < z = -1.645$, reject H_0.

Conclude that the brick facings made by plant 2 have a larger mean thickness than those made by plant 1.

Hypothesis-Testing for Mu Sub 1 Minus Mu Sub 2 when Sigma Sub 1 and Sigma Sub 2 Are Known-Example *p*-Value Example

If *p*-value < α, reject the null hypothesis; Otherwise, do not reject the null hypothesis.

Example:

- The *p*-value for one-tailed test is the probability of a *z*-value in a standard normal distribution being less than −1.90.
- The probability associated with $z = -1.90$ is 0.4713
- *p*-value = 0.5000 − 0.4713 = 0.0287
- *p*-value $= 0.0287 < \alpha = 0.05$
- Reject null hypothesis

Testing for Mu Sub 1 Minus Mu Sub 2 When Sigma Sub 1 and Sigma Sub 2 Are Unknown Independent Samples (1 of 3)

- *t*-Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

$\bar{x}_1$ and $\bar{x}_2$ - Sample means from population 1 and 2

$\mu_1 - \mu_2$ - Hypothesized difference between population means

n_1 and n_2 - Sample sizes from two populations

s_p - Pooled standard deviation

Testing for Mu Sub 1 Minus Mu Sub 2 When Sigma Sub 1 and Sigma Sub 2 Are Unknown Independent Samples (2 of 3)

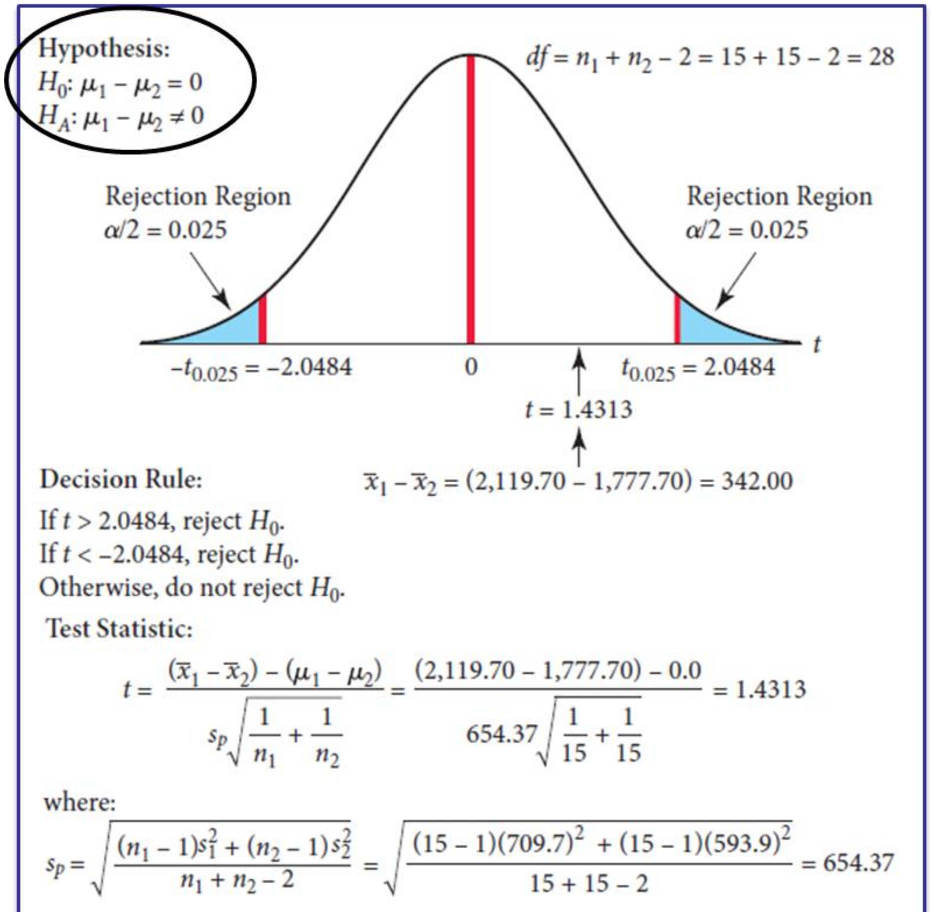
One of the big costs associated with using computer printers is the cost of ink cartridges. An independent testing laboratory is interested in testing to see whether the mean number of pages generated per cartridge is the same for the popular name-brand as it is for the leading generic brand. To conduct the test at an alpha = 0.05 level, the lab has randomly selected 15 cartridges of each type and counted the number of pages generated by each. The following sample data were observed:

$$n_1 = 15 \quad n_2 = 15$$

$$\bar{x}_1 = 2{,}119.7 \quad \bar{x}_2 = 1{,}777.7$$

$$s_1 = 709.7 \quad s_2 = 593.9$$

Testing for Mu Sub 1 Minus Mu Sub 2 When Sigma Sub 1 and Sigma Sub 2 Are Unknown Independent Samples (3 of 3)



Because $t = 1.4313 < t_{0.025} = 2.0484$, do not reject the null hypothesis. Conclude that there may be no difference in population means.

The Hypothesis-Testing Process for Two Population Means - Example

One-tailed (upper) Test Example:

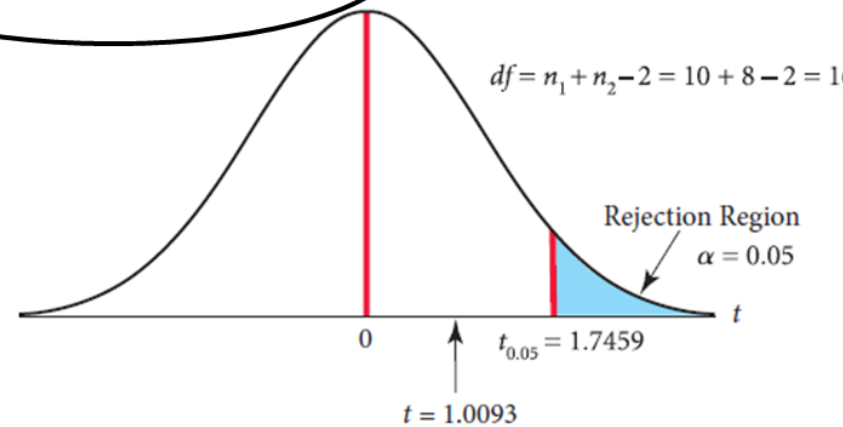
The leaders of the study are interested in determining whether there is a difference in mean annual contributions for individuals covered by TSAs and those with 401(k) retirement programs.

TSA	401(k)
$n_1 = 10$	$n_2 = 8$
$\bar{x}_1 = 322.5$	$\bar{x}_2 = 298.3$
$s_1 = 48.3$	$s_2 = 53.3$

$H_0: \mu_1 - \mu_2 \le 0$ or $H_0: \mu_1 \le \mu_2$

$H_A: \mu_1 - \mu_2 > 0$ $H_A: \mu_1 > \mu_2$

$\alpha = 0.05$



where:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)\,48.3^2 + (8-1)\,53.3^2}{10+8-2}} = 50.55$$

Test Statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(322.5 - 298.3) - 0}{50.55\sqrt{\frac{1}{10} + \frac{1}{8}}} = 1.0093$$

Decision Rule:

Because $t = 1.0093 < t_{0.05} = 1.7459$, do not reject H_0.

Hypothesis Test for Two Populations – Example - How to Do It in Excel

The Environmental Protection Agency (EPA) wishes to run a test on a new Hybrid vehicle to see whether there is difference between the mean highway mpg and the mean city mpg are equal. The EP{A technicians have run a test involving a random sample of 25 hybrid vehicles driven on the highway only and another 25 of the hybrid vehicles driven in the city only. The sample data are stored in an Excel file.

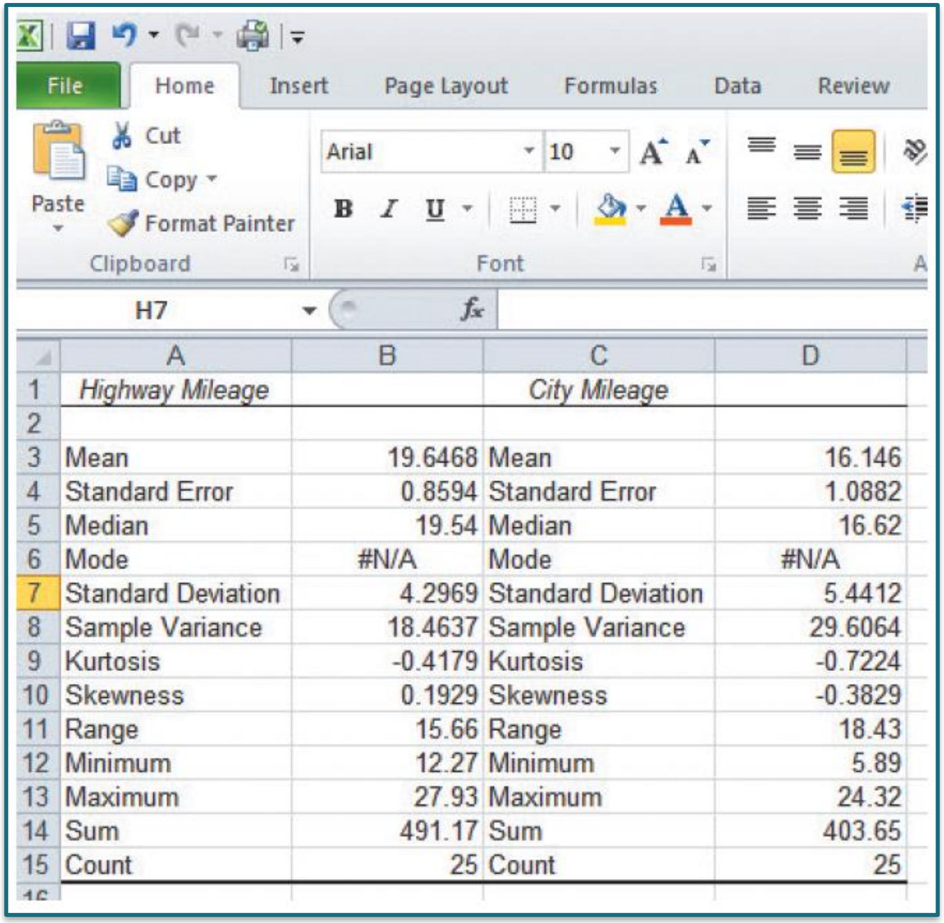
$$H_o : \mu_1 - \mu_2 = 0.0$$

$$H_A : \mu_1 - \mu_2 \neq 0.0$$

$$\alpha = 0.05$$

Step 1: Use the Descriptive Statistics Tool - Optional

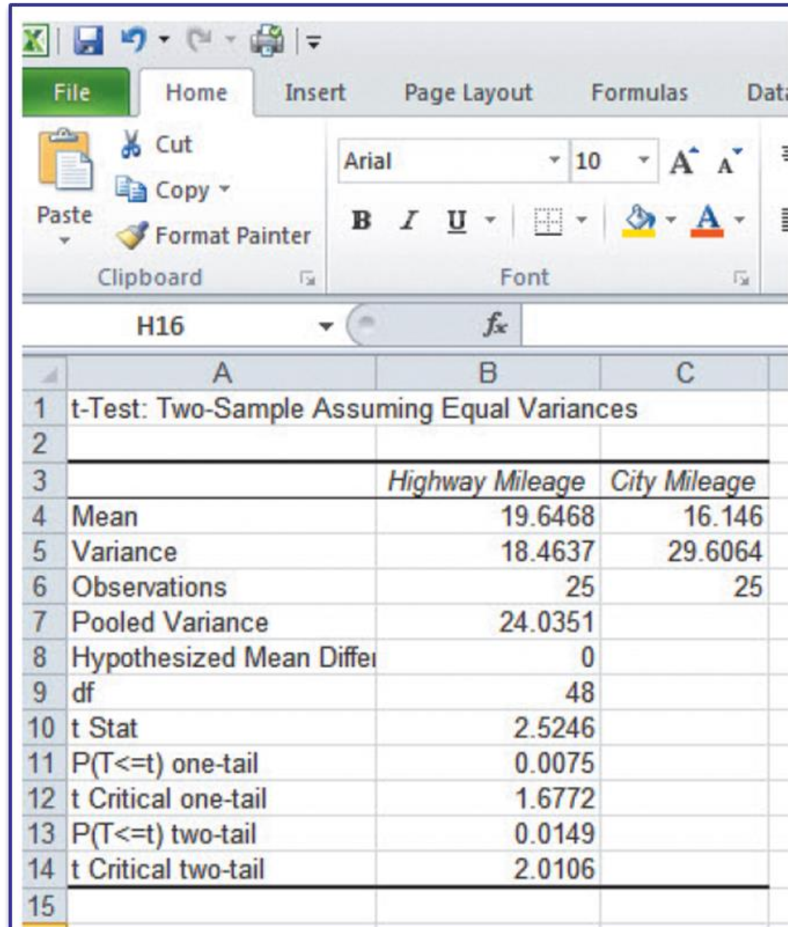
1. Open file.
2. Select **Data** > **Data Analysis**
3. Select **Descriptive Statistics**.
4. Define the data range for all variables to be analyzed.
5. Select **Summary Statistics**.
6. Specify output location.
7. Click **OK.**



	A	B	C	D
1	*Highway Mileage*		*City Mileage*	
2				
3	Mean	19.6468	Mean	16.146
4	Standard Error	0.8594	Standard Error	1.0882
5	Median	19.54	Median	16.62
6	Mode	#N/A	Mode	#N/A
7	Standard Deviation	4.2969	Standard Deviation	5.4412
8	Sample Variance	18.4637	Sample Variance	29.6064
9	Kurtosis	-0.4179	Kurtosis	-0.7224
10	Skewness	0.1929	Skewness	-0.3829
11	Range	15.66	Range	18.43
12	Minimum	12.27	Minimum	5.89
13	Maximum	27.93	Maximum	24.32
14	Sum	491.17	Sum	403.65
15	Count	25	Count	25

Step 2: Use the Data Analysis Tool: *t*-Test: Two Sample Assuming Equal Variances

1. Open file.
2. Select **Data** > **Data Analysis.**
3. Select ***t*-test: Two Sample Assuming Equal Variances**.
4. Define data ranges for the two variables of interest.
5. Set **Hypothesized Difference** to 0.0.
6. Set **Alpha** value.
7. Specify output location.
8. Click **OK**.
9. Click the **Home** tab and adjust decimal places in output.



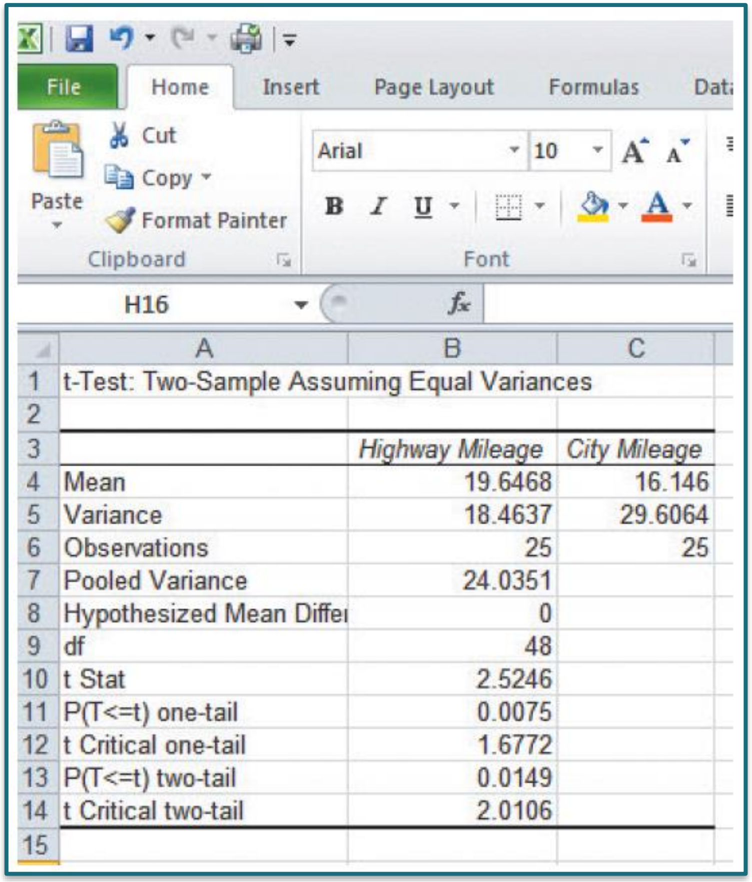
	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Highway Mileage*	*City Mileage*
4	Mean	19.6468	16.146
5	Variance	18.4637	29.6064
6	Observations	25	25
7	Pooled Variance	24.0351	
8	Hypothesized Mean Differ	0	
9	df	48	
10	t Stat	2.5246	
11	P(T<=t) one-tail	0.0075	
12	t Critical one-tail	1.6772	
13	P(T<=t) two-tail	0.0149	
14	t Critical two-tail	2.0106	
15			

Step 3: Analyze the Excel Output

$$H_o: \mu_1 - \mu_2 = 0.0$$

$$H_A: \mu_1 - \mu_2 \neq 0.0$$

$$\alpha = 0.05$$



	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Highway Mileage*	*City Mileage*
4	Mean	19.6468	16.146
5	Variance	18.4637	29.6064
6	Observations	25	25
7	Pooled Variance	24.0351	
8	Hypothesized Mean Differ	0	
9	df	48	
10	t Stat	2.5246	
11	P(T<=t) one-tail	0.0075	
12	t Critical one-tail	1.6772	
13	P(T<=t) two-tail	0.0149	
14	t Critical two-tail	2.0106	
15			

Because *t* Stat $= 2.5246 > t_{0.025,\mathrm{df}=48} = 2.0106$, we reject the null hypothesis and conclude that the vehicle gets significantly higher mpg in highway driving.

Testing for Mu Sub 1 Minus Mu Sub 2 When Sigma Sub 1 and Sigma Sub 2 Are Known and Not Equal

- *t*-Test Statistic:

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

- Degrees of Freedom:

$$df = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\left(\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)}$$

10.3 Interval Estimation and Hypothesis Tests for Paired Samples

- Paired samples are dependent samples
- Samples that are selected in such a way that values in one sample are matched with the values in the second sample for the purpose of controlling for extraneous factors
- Examples
  - Testing a new paint mix vs an old one
  - Testing difference in gas mileage comparing regular and premium gas

Interval Estimation and Hypothesis Tests for Paired Samples (1 of 2)

- Paired Difference:

$$d = x_1 - x_2$$

x_1 and x_2 – Values from samples 1 and 2, respectively

- Point Estimate for the Population Mean Paired Difference μ_d :

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

d_i - i^{th} paired difference

$\bar{d}$ - Mean paired difference

n - Number of pairs

Interval Estimation and Hypothesis Tests for Paired Samples (2 of 2)

- Sample Standard Deviation for Paired Differences Samples:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n}}$$

d_i - i^{th} paired difference

$\bar{d}$ - Mean paired difference

n - Number of pairs

- Confidence Interval Estimate for Population Mean Paired Difference, μ_d :

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

t - Critical *t* value from *t*-distribution with *n* − 1 degrees of freedom

s_d - Sample standard deviation of paired differences

Interval Estimation and Hypothesis Tests for Paired Samples - Example (1 of 2)

A consumer magazine wishes to estimate the difference in mean mpg for vehicles that use ethanol fuel versus non-ethanol fuel. Ten vehicles were selected randomly and each vehicle was tested using ethanol and non-ethanol fuel Each car was driven 200 miles on an identical route. Develop a 95% confidence interval estimate for the mean paired difference in mpg resulting from the two fuels. The sample data and paired differences are shown as follows:

	A	B	C
1	Ethanol	Regular	d
2	19.8	20.7	-0.9
3	28.8	25.8	3
4	20.4	27.8	-7.4
5	18.7	14.9	3.8
6	23.4	21.6	1.8
7	27.1	21.1	6
8	28.4	28	0.4
9	21.4	13	8.4
10	26.4	24.4	2
11	19.9	14.3	5.6

Interval Estimation and Hypothesis Tests for Paired Samples - Example (2 of 2)

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{22.7}{10} = 2.27$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{172.8}{10-1}} = 4.38$$

$$\bar{d} \pm t_{0.95, df=9} \frac{s_d}{\sqrt{n}}$$

$$2.27 \pm 2.2622 \frac{4.38}{\sqrt{10}}$$

$$2.27 \pm 3.13$$

–0.86 mpg -------- 5.40 mpg

Hypothesis Testing for Paired Samples (1 of 2)

- *t*-Test Statistic for Paired-Sample Test:

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

$\bar{d}$ - Mean paired difference

μ_d - Hypothesized population mean paired difference

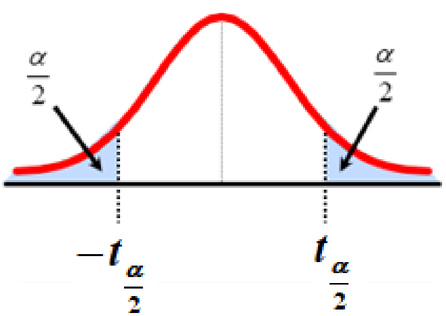
s_d - Sample standard deviation for paired differences

n - Number of paired values in the sample

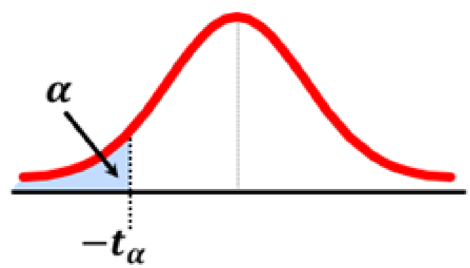
Hypothesis Testing for Paired Samples (2 of 2)

Variations in Hypothesis Testing

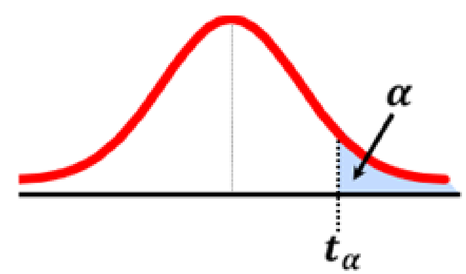
Two-tailed Test	One-tailed Test (lower)	One-tailed Test (upper)
$H_0: \mu_d = 0$	$H_0: \mu_d \geq 0$	$H_0: \mu_d \leq 0$
$H_A: \mu_d \neq 0$	$H_A: \mu_d < 0$	$H_A: \mu_d > 0$



Reject H_0 if $t < -t_{\frac{\alpha}{2}}$ or $t > t_{\frac{\alpha}{2}}$



Reject H_0 if $t < -t_{\alpha}$



Reject H_0 if $t > t_{\alpha}$

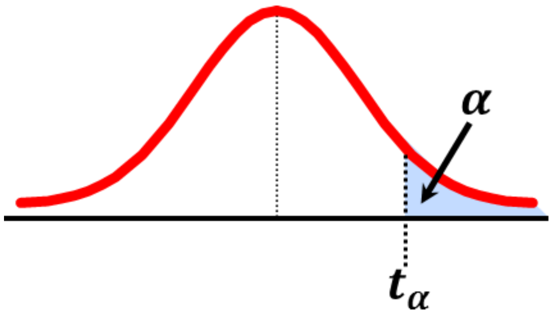
Hypothesis Testing for Paired Samples - Example

Suppose an independent test agency wishes to conduct a test to determine whether name-brand ink cartridges generate more color pages on average than competing generic ink cartridges. The test is conducted using paired samples. This means that the same people will use both types of cartridges, and the pages printed in each case will be recorded.

$$H_0 : \mu_d \leq 0$$

$$H_A : \mu_d > 0$$

$$\alpha = 0.01 \qquad n = 6 \qquad df = 5 \qquad t_\alpha = 3.3649$$



Reject H_0 **if** $t > t_\alpha$

Hypothesis Testing for Paired Samples - Solution

User	Name-Brand	Generic	d_i
1	306	300	6
2	256	260	−4
3	402	357	45
4	299	286	13
5	306	290	16
6	257	260	−3

- The mean paired difference: $\bar{d} = \frac{73}{6} = 12.17$
- The standard deviation for the paired differences: $s_d = 18.02$
- The *t*-test statistic: $t = \frac{12.17 - 0.0}{\frac{18.02}{\sqrt{6}}} = 1.6543$
- Decision and conclusion: $t = 1.6543 < t_{0.01} = 3.3649$:

Do not reject the null hypothesis

10.4 Estimation and Hypothesis Tests for Two Population Proportions

- Assumptions: $n_1 \bar{p}_1 \geq 5; \quad n_1 \left(1 - \bar{p}_1\right) \geq 5$

  $n_2 \bar{p}_2 \geq 5; \quad n_2 \left(1 - \bar{p}_2\right) \geq 5$

- Point Estimate for $p_1 - p_2$: $\bar{p}_1 - \bar{p}_2$
- Confidence Interval Estimate for $p_1 - p_2$:

$$\left(\bar{p}_1 - \bar{p}_2\right) \pm z \sqrt{\frac{\bar{p}_1 \left(1 - \bar{p}_1\right)}{n_1} + \frac{\bar{p}_2 \left(1 - \bar{p}_2\right)}{n_2}}$$

$\bar{p}_1$ - Sample proportion from population 1

$\bar{p}_2$ - Sample proportion from population 2

z - Critical value from the standard normal table

Estimation for Two Population Proportions - Example (1 of 2)

An outdoor sports publication wishes estimate the difference between male and female bicycle enthusiasts in terms of how they rate the quality of new high-end mountain bike. A random sample of 425 men and 370 women were asked to rate the bike and the editors counted the number of each gender who rated the bike as "very high" quality. The following shows the results:

Men	Women
$n_1 = 425$	$n_2 = 370$
$x_1 = 240$	$x_2 = 196$

$$\bar{p}_1 = \frac{240}{425} = 0.565 \quad \bar{p}_2 = \frac{196}{370} = 0.530$$

Estimation for Two Population Proportions - Example (2 of 2)

$$\text{Point Estimate} \pm (\text{Critical Value})(\text{Standard Error})$$

$$(\bar{p}_1 - \bar{p}_2) \pm z_{0.95}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

$$(0.565 - 0.530) \pm 1.96\sqrt{\frac{0.565(1-0.565)}{425} + \frac{0.530(1-0.535)}{370}}$$

$$0.035 \pm 0.069$$

$$-0.034 \text{ -------------- } 0.104$$

Hypothesis Testing for Two Population Proportions

Variations in Hypothesis Testing

Two-tailed Test

$H_0 : p_1 - p_2 = 0.0$

$H_A : p_1 - p_2 \neq 0.0$

or

$H_0 : p_1 = p_2$

$H_A : p_1 \neq p_2$

One-tailed Test (lower)

$H_0 : p_1 - p_2 \geq 0.0$

$H_A : p_1 - p_2 < 0.0$

or

$H_0 : p_1 \geq p_2$

$H_A : p_1 < p_2$

One-tailed Test (upper)

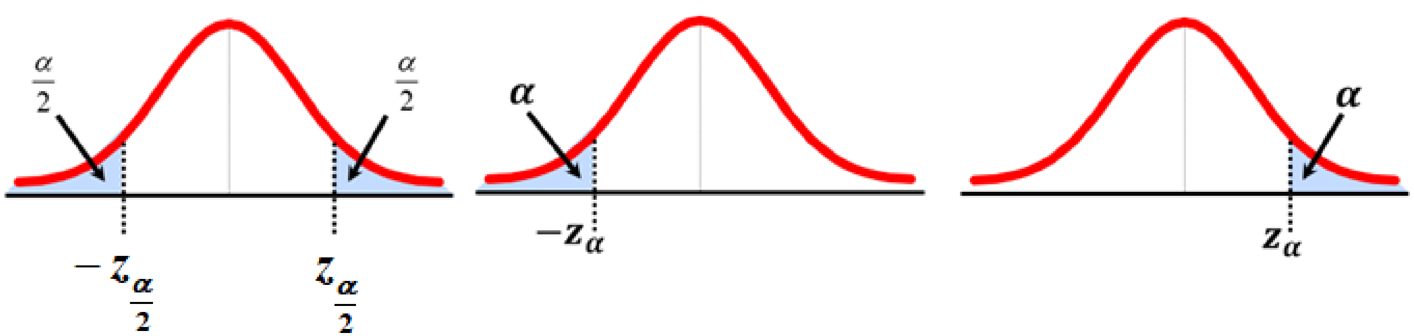
$H_0 : p_1 - p_2 \leq 0.0$

$H_A : p_1 - p_2 < 0.0$

or

$H_0 : p_1 \leq p_2$

$H_A : p_1 < p_2$



Reject H_0 **if** $z < -z_{\frac{\alpha}{2}}$ **or** $z > z_{\frac{\alpha}{2}}$

Reject H_0 **if** $z < -z_{\alpha}$

Reject H_0 **if** $z > z_{\alpha}$

Hypothesis Tests for Two Population Proportions

- Pooled Estimator for Overall Proportion:

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

x_1 and x_2 - Number from samples 1 and 2 with the characteristic of interest

- *z*-Test Statistic for Difference between Population Proportions:

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

$\left(p_1 - p_2\right)$ - Hypothesized difference in proportions from two populations

$\bar{p}_1$ and $\bar{p}_2$ - Sample proportions for samples selected from populations

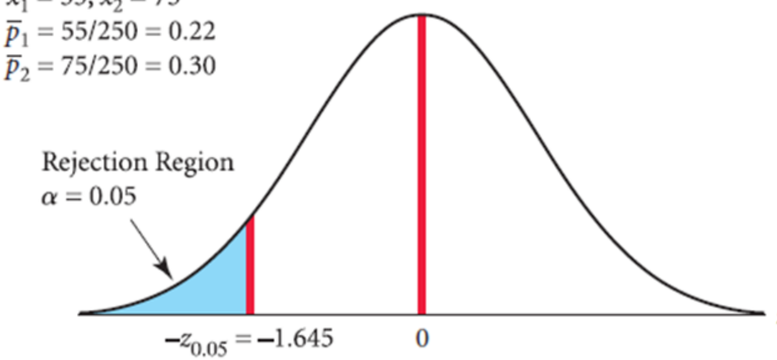
$\bar{p}$ - Pooled estimator for the overall proportion for both populations combined

Hypothesis Testing for Two Population Proportions - Example

A critical component of a handheld hair dryer is the motor-heater unit. Company has recently created a new motor-heater unit with fewer parts than the current unit. Company has decided to test samples of old and new units to see which motor-heater is more reliable. The null hypothesis states that the new motor-heater is no better than the old, or current, motor-heater.

New Unit	Old Unit
$n_1 = 250$	$n_2 = 250$
$x_1 = 55$	$x_2 = 75$
$\bar{p}_1 = 0.22$	$\bar{p}_2 = 0.30$

H_0: $p_1 - p_2 \geq 0$
H_A: $p_1 - p_2 < 0$
$n_1 = 250, n_2 = 250$
$x_1 = 55, x_2 = 75$
$\bar{p}_1 = 55/250 = 0.22$
$\bar{p}_2 = 75/250 = 0.30$



Decision Rule:

If $z < -1.645$, reject H_0.

If $z \geq -1.645$, do not reject H_0.

Test Statistic:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.22 - 0.30) - 0}{\sqrt{0.26(1-0.26)\left(\frac{1}{250} + \frac{1}{250}\right)}} = -2.04$$

where:

$$\bar{p} = \frac{250(0.22) + 250(0.30)}{250 + 250} = \frac{55 + 75}{500} = 0.26$$

Since $z = -2.04 < -1.645$, reject H_0.