Dr. Saeed Alghamdi
Room 542
Statistics Department
Faculty of Sciences
King Abdulaziz University
web sit
http://saalghamdy.kau.edu.sa

Elementary Statistics

A Step by Step Approach
Eighth Edition

by
Allan G. Bluman

SLIDES PREPARED
BY

Dr. Saeed A Dobbah Alghamdi
Statistics Department
King Abdulaziz University

CHAPTER 9

Testing the Difference Between Two Means, Two Proportions, and Two Variances

Notes

❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………

Objectives

- Test the difference between two sample means using the *z* test.
- Test the difference between two means for independent samples using the *t* test.
- Test the difference between two means for dependent samples using the *t* test.
- Test the difference between two proportions.

Notes

❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………
❑……………………………………………………………

Introduction

- There are many instances when researchers wish to compare two sample means using experimental and control groups.
- For example, two brands of cough syrup might be tested to see whether one brand is more effective than the other.
- In the comparison of two means or proportions, the same basic steps for hypothesis testing shown before are used.
- The *z* and *t* tests are also used.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *z* Test

The hypotheses are

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \neq \mu_2$ *Two − tailed*

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 > \mu_2$ *Right − tailed*

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 < \mu_2$ *Left − tailed*

Assumptions:

1. The samples must be independent of each other. Thus, there is no relationship between the subjects in each sample.
2. The standard deviations of both populations must be known, and if the sample size is less than 30, the populations must be normally or approximately normally distributed.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *z* Test

The formula for comparing two means from independent populations is as the following:

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

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

σ_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *z* Test

The formula for the z confidence interval for difference between tow means are as the following:

$$\left(\bar{X}_1 - \bar{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{X}_1 - \bar{X}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

σ_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

Notes

- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………

A survey of 50 hotels in each city found that the average hotel room rate in Jeddah is 400 riyals and the average room rate in Riyadh is 350 riyals. The standard deviations of the populations are 56 and 48 riyals respectively. At $\alpha = 0.01$, can it be concluded that there is a significant difference in the rates?

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

Since $\alpha = 0.01$ and the test is a two-tailed test, the critical values are -2.58 and 2.58.

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(400 - 350) - 0}{\sqrt{\frac{56^2}{50} + \frac{48^2}{50}}} = 4.79$$

Since the value of z is greater than 2.58, reject the null hypothesis. Thus, there is enough evidence to support the claim that the average hotel room rates is different in the two cities. Note that the confidence interval for the difference between the two means is

$$(400 - 350) - 2.58 \times \sqrt{\frac{56^2}{50} + \frac{48^2}{50}} < \mu_1 - \mu_2 < (400 - 350) + 2.58 \times \sqrt{\frac{56^2}{50} + \frac{48^2}{50}} \Rightarrow 23.09 < \mu_1 - \mu_2 < 76.91$$

Notes

- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………

A researcher hypothesize that the average number of sports that colleges offer for males is greater than the average number of sports that colleges offer for females. A sample of the number of sports offered by colleges is in the attached file. At $\alpha = 0.10$, is there enough evidence to support the claim? Assume σ_1 and $\sigma_2 = 3.3$.

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 > \mu_2$

1. From the toolbar, select Add-Ins, MegaStat>Hypothesis Tests>Compare Two Independent Groups.
2. Select data input.
3. Select the data for group 1 and group 2.
4. Select the “greater than” Alternative.
5. Select z-test.
6. Select the 90% confidence interval.
7. Click [OK]

Notes

- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………
- ………………………………………………………………

A researcher hypothesize that the average number of sports that colleges offer for males is greater than the average number of sports that colleges offer for females. A sample of the number of sports offered by colleges is in the attached file. At $\alpha = 0.10$, is there enough evidence to support the claim? Assume σ_1 and $\sigma_2 = 3.3$.

Descriptive statistics

		Male	Female
		8.56	7.94
	Males		
count	50	3.3	3.3
mean	8.56	50	50
	Females		
count	50		
mean	7.94		

Hypothesis Test: Independent Groups (z-test)

Male	Female	
8.56	7.94	mean
3.3	3.3	std. dev.
50	50	n

0.62000	difference (Male - Female)
0.66000	standard error of difference
0	hypothesized difference
0.94	z
.1738	p-value (one-tailed, upper)
-0.46560	confidence interval 90.% lower
1.70560	confidence interval 90.% upper
1.08560	margin of error

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 > \mu_2$

Since the p-value is 0.1738 which is greater than $\alpha = 0.10$, then do not reject the null hypothesis. Thus, there is not enough evidence to support the claim that colleges offer more sports for males than they do for females. Note that the confidence interval contains 0.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *t* Test

- In many situations the population standard deviations are not known, hence a ***t* test** is used to test the difference between means.
- The samples can be ***independent*** in which they are not related or ***dependent*** in which they are paired or matched in some way, hence there are different formulas for using the ***t* test** for testing the difference between two means.
- Depending on the situation, we will have ***t* test** formula for ***independent samples*** assuming different populations variances and another one assuming equal populations variances. Also, there is another formula for ***dependent samples***.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *t* Test

- To use the *t* test in the case of independent populations, first determine whether the variances of the two populations are equal. Then use the appropriate *t* test formula.
- The pooled estimate of variance is used to calculate the standard error in the *t* test when the variances are equal.
- A *pooled estimate of the variance* is a weighted average of the variance using the two sample variances and the degrees of freedom of each variance as the weights.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

The Difference Between Two Means Using the *t* Test

The formula for comparing two means from independent populations assuming unequal populations variances is as the following:

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

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

s_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

the degrees of freedom are equal to the smaller of $n_1 - 1$ or $n_2 - 1$

Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………

The Difference Between Two Means Using the *t* Test

The formula for the *t* confidence interval for difference between tow means from independent populations assuming unequal populations variances is as the following:

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

s_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

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

Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………

In a statistics exam, the average grade for students who had form A is 18 with standard deviation 3.34 for a sample size of 29 students and the average grade for students who had form B is 16.76 with standard deviation 5.46 for a sample size of 29 students. Can it be concluded at $\alpha = 0.05$ that the average grades for the two forms is different? Assume the populations are normally distributed.

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

1. From the toolbar, select Add-Ins, MegaStat>Hypothesis Tests>Compare Two Independent Groups.
2. Select summary input.
3. Select the data for group 1 and group 2.
4. Select the "not equal" Alternative.
5. Select the 95% confidence interval.
6. Check the "Test for equality of variances" box.
7. Select the appropriate t-test (pooled or unequal variance)
8. Click [OK]

Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………

In a statistics exam, the average grade for students who had form A is 18 with standard deviation 3.34 for a sample size of 29 students and the average grade for students who had form B is 16.76 with standard deviation 5.46 for a sample size of 29 students. Can it be concluded at $\alpha = 0.05$ that the average grades for the two forms is different? Assume the populations are normally distributed.

$$H_0: \mu_1 = \mu_2 \quad \text{and} \quad H_1: \mu_1 \neq \mu_2$$

Hypothesis Test: Independent Groups (t-test, unequal variance)

Form A	Form B	
18	16.76	mean
3.34	5.46	std. dev.
29	29	n
	46	df
	1.24000	difference (Form A - Form B)
	1.18855	standard error of difference
	0	hypothesized difference
	1.04	t
	.3023	p-value (two-tailed)
	-1.15244	confidence interval 95.% lower
	3.63244	confidence interval 95.% upper
	2.39244	margin of error
F-test for equality of variance		
	29.8116	variance: Form B
	11.1556	variance: Form A
	2.67	F
	.0114	p-value

Since the p-value is 0.3023 is greater than $\alpha = 0.05$, do not reject the null hypothesis. Thus, there is not enough evidence to support the claim that the average grades is different. Note that the confidence interval contains 0.

Notes

❑..

❑..

❑..

❑..

❑..

❑..

The Difference Between Two Means Using the *t* Test

The formula for comparing two means from independent populations assuming <u>equal populations variances</u> is as the following:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

s_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

the degrees of freedom are equal to $n_1 + n_2 - 2$

Notes

❑..

❑..

❑..

❑..

❑..

❑..

The Difference Between Two Means Using the *t* Test

The formula for the *t* confidence interval for difference between tow means from independent populations assuming <u>equal populations variances</u> is as the following:

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$< \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$\bar{X}_i$ = sample mean from population i=1, 2

μ_i = hypothesized population mean; i=1, 2

s_i^2 = population variance; i=1, 2

n_i = sample size; i=1, 2

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

Notes

❑..

❑..

❑..

❑..

❑..

❑..

In a statistics exam, the average grade for students who had form A is 15.81 with standard deviation 5.54 for a sample size of 28 students and the average grade for students who had form B is 16.76 with standard deviation 5.46 for a sample size of 29 students. Can it be concluded at $\alpha = 0.05$ that the average grades for the two forms is different? Assume the populations are normally distributed.

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad H_1 : \mu_1 \neq \mu_2$$

1. From the toolbar, select Add-Ins,
   MegaStat>Hypothesis Tests>Compare Two Independent Groups.
2. Select summary input.
3. Select the data for group 1 and group 2.
4. Select the "not equal" Alternative.
5. Select the 95% confidence interval.
6. Check the "Test for equality of variances" box.
7. Select the appropriate t-test (pooled or unequal variance)
8. Click [OK]

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

In a statistics exam, the average grade for students who had form A is 15.81 with standard deviation 5.54 for a sample size of 28 students and the average grade for students who had form B is 16.76 with standard deviation 5.46 for a sample size of 29 students. Can it be concluded at $\alpha = 0.05$ that the average grades for the two forms is different? Assume the populations are normally distributed.

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad H_1 : \mu_1 \neq \mu_2$$

Hypothesis Test: Independent Groups (t-test, pooled variance)

Form A	Form B	
15.81	16.76	mean
5.54	5.46	std. dev.
28	29	n
	55	df
	-0.95000	difference (Form A - Form B)
	30.24360	pooled variance
	5.49942	pooled std. dev.
	1.45706	standard error of difference
	0	hypothesized difference
	-0.65	t
	.5171	p-value (two-tailed)
	-3.87000	confidence interval 95.% lower
	1.97000	confidence interval 95.% upper
	2.92000	margin of error
F-test for equality of variance		
	30.6916	variance: Form A
	29.8116	variance: Form B
	1.03	F
	.9379	p-value

Since the p-value is 0.5171 is greater than $\alpha = 0.05$, do not reject the null hypothesis. Thus, there is not enough evidence to support the claim that the average grades is different. Note that the confidence interval contains 0.

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

The Difference Between Two Means Using the *t* Test

- Samples in which the same subjects are used in a pre-post situation are dependent.
- Another type of dependent samples are samples matched on the basis of variables extraneous to the study.

Caution:

1. When subjects are matched according to one variable, the matching process does not eliminate the influence of other variables.
2. When the same subjects are used for a pre-post study, sometimes the knowledge that they are participating in a study can influence the results.

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

The Difference Between Two Means Using the *t* Test

- Hypotheses:

Two-tailed	Left-tailed	Right-tailed
H_0: $\mu_D = 0$	H_0: $\mu_D = 0$	H_0: $\mu_D = 0$
H_1: $\mu_D \neq 0$	H_1: $\mu_D < 0$	H_1: $\mu_D > 0$

- μ_D is the expected mean of the differences of the matched pairs.

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

The Difference Between Two Means Using the *t* Test

General Procedure for Finding the Test Value

Step 1 Find the differences of the values of the pairs of data, *D*.

Step 2 Find the mean of the differences .

Step 3 Find the standard deviation of the differences, .

Step 4 Find the estimated standard error of the differences, .

Step 5 Find the test value, *t*.

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

The Difference Between Two Means Using the *t* Test

- The formula for the *t* test for dependent samples:

$$t = \frac{\overline{D} - \mu_D}{s_D / \sqrt{n}}$$

with d.f. = *n* – 1 and

where
$$\overline{D} = \frac{\Sigma D}{n}$$

and
$$s_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{n-1}}$$

Notes

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

❑…………………………………………………………………

The Difference Between Two Means Using the *t* Test

- The formula for calculating the confidence interval for the mean difference:

$$\overline{D} - t_{\alpha/2}\frac{s_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\alpha/2}\frac{s_D}{\sqrt{n}}$$

d.f. $= n - 1$

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

- Test the hypothesis that there is no difference in population means, based on these sample paired data. Use $\alpha = 0.05$.

Set A	33	35	28	29	32	34	30	34
Set B	27	29	36	34	30	29	28	24

$H_0: \mu_D = 0$ and $H_1: \mu_D \neq 0$

1. From the toolbar, select Add-Ins, MegaStat>Hypothesis Tests>Paired Observations.
2. Select summary input.
3. Select the data for group 1 and group 2.
4. Enter the value 0 for Hypothesized difference.
5. Select the “not equal” Alternative.
6. Select t-test.
7. Select the 95% confidence interval.
8. Click [OK]

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

- Test the hypothesis that there is no difference in population means, based on these sample paired data. Use $\alpha = 0.05$.

Set A	33	35	28	29	32	34	30	34
Set B	27	29	36	34	30	29	28	24

$H_0: \mu_D = 0$ and $H_1: \mu_D \neq 0$

Hypothesis Test: Paired Observations

0.000	hypothesized value
31.875	mean Set A
29.625	mean Set B
2.250	mean difference (Set A - Set B)
6.018	std. dev.
2.128	std. error
8	n
7	df
1.06	t
.3254	p-value (two-tailed)
-2.781	confidence interval 95.% lower
7.281	confidence interval 95.% upper
5.031	margin of error

Since the p-value 0.3254 is greater than 0.05, do not reject the null hypothesis. Thus, there is enough evidence to conclude that there is no difference in population means. Note that the confidence interval contains 0.

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

Testing the Difference Between Proportions

- $\hat{p}$,"p hat", is the sample proportion that is used to estimate the population proportion.

$$\hat{p} = \frac{X}{n}$$

where

X = number of units that possess the characteristic of interest

n = sample size

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

Testing the Difference Between Proportions

- For population proportions, p_1 and p_2 the hypotheses can be stated as follows, if no difference between the proportions is hypothesized.
- H_0: $p_1 = p_2$ $\qquad$ H_0: $p_1 - p_2 = 0$
- H_1: $p_1 \neq p_2$ $\qquad$ H_1: $p_1 - p_2 \neq 0$
- $\hat{p}_1 = X_1 / n_1$ is used to estimate p_1.
- $\hat{p}_2 = X_2 / n_2$ is used to estimate p_2.

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

Testing the Difference Between Proportions

- Since p_1 and p_2 are unknown, a weighted estimate of p can be computed by using the formula below.

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}$$

- The standard error of the difference in terms of the weighted estimate is:

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

where

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} \qquad \hat{p}_1 = X_1 / n_1$$

$$\bar{q} = 1 - \bar{p} \qquad \hat{p}_2 = X_2 / n_2$$

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

Testing the Difference Between Proportions

- The formula for the *z* test for comparing two proportions:

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

- The confidence interval for the difference between two proportions can be calculated using the following formula:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$< p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

In a sample of 200 workers, 45% said that they missed work because of personal illness. Ten years ago in a sample of 200 workers, 35% said that they missed work because of personal illness. At $\alpha = 0.01$, is there a difference in the proportions?

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

1. From the toolbar, select Add-Ins, MegaStat>Hypothesis Tests>Compare Two Proportions.
2. Enter the data for group 1 and group 2.
3. Enter the value 0 for Hypothesized difference.
4. Select the “not equal” Alternative.
5. Select the 99% confidence interval.
6. Click [OK]

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

In a sample of 200 workers, 45% said that they missed work because of personal illness. Ten years ago in a sample of 200 workers, 35% said that they missed work because of personal illness. At $\alpha = 0.01$, is there a difference in the proportions?

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

Hypothesis test for two independent proportions

p1	p2	pc	
0.45	0.35	0.4	p (as decimal)
90/200	70/200	160/400	p (as fraction)
90.	70.	160.	X
200	200	400	n

0.1	difference
0.	hypothesized difference
0.049	std. error
2.04	z
.0412	p-value (two-tailed)
-0.0255	confidence interval 99.% lower
0.2255	confidence interval 99.% upper
0.1255	margin of error

Since the p-value 0.0412 is greater than $\alpha = 0.01$, do not reject the null hypothesis. Thus, there is not enough evidence to support the claim that there is a difference in proportions. Note that the confidence interval contains 0.

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

Summary

- Means and Proportions are population parameters that are often compared.
- This comparison can be made with the z test if the samples are independent and the population variances are known.
- If the variances are not known for at least one population, then the t test must be used.
- For dependent samples the dependent samples t test is used.
- A z test is used to compare two proportions.

Notes

- ❑……………………………………………………………………
- ❑……………………………………………………………………
- ❑……………………………………………………………………
- ❑……………………………………………………………………
- ❑……………………………………………………………………
- ❑……………………………………………………………………