Business Statistics: A Decision-Making Approach

6th Edition

Chapter 9

Estimation and Hypothesis Testing for Two Population Parameters

Chapter Goals

After completing this chapter, you should be able to:

- Test hypotheses or form interval estimates for
  - two independent population means
    - Standard deviations known
    - Standard deviations unknown
  - two means from paired samples
  - the difference between two population proportions

Estimation for Two Populations

Estimating two population values

- Population means, independent samples
- Paired samples
- Population proportions

Examples:

- Group 1 vs. independent Group 2
- Same group before vs. after treatment
- Proportion 1 vs. Proportion 2

Difference Between Two Means

- Population means, independent samples *
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

Goal: Form a confidence interval for the difference between two population means, $\mu_1 - \mu_2$

The point estimate for the difference is

$$\bar{x}_1 - \bar{x}_2$$

 Chap 9-4

Independent Samples

- Population means, independent samples *
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

- Different data sources
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
- Use the difference between 2 sample means
- Use z test or pooled variance *t* test

 Chap 9-5

σ_1 and σ_2 known

- Population means, independent samples
  - σ_1 and σ_2 known *
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

Assumptions:

- Samples are randomly and independently drawn
- population distributions are normal or both sample sizes are ≥ 30
- Population standard deviations are known

 Chap 9-6

σ_1 and σ_2 known

(continued)

- Population means, independent samples
  - σ_1 and σ_2 known *
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

When σ_1 and σ_2 are known and both populations are normal or both sample sizes are at least 30, the test statistic is a z-value…

…and the standard error of $\bar{x}_1 - \bar{x}_2$ is

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-7

σ_1 and σ_2 known

(continued)

- Population means, independent samples
  - σ_1 and σ_2 known *
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

The confidence interval for $\mu_1 - \mu_2$ is:

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-8

σ_1 and σ_2 unknown, large samples

- Population means, independent samples
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$ *
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

Assumptions:

- Samples are randomly and independently drawn
- both sample sizes are ≥ 30
- Population standard deviations are unknown

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-9

σ_1 and σ_2 unknown, large samples

(continued)

- Population means, independent samples
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$ *
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

Forming interval estimates:

- use sample standard deviation s to estimate σ
- the test statistic is a z value

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-10

σ_1 and σ_2 unknown, large samples

(continued)

- Population means, independent samples
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$ *
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

The confidence interval for $\mu_1 - \mu_2$ is:

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-11

σ_1 and σ_2 unknown, small samples

- Population means, independent samples
  - σ_1 and σ_2 known
  - σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
  - σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ *

Assumptions:

- populations are normally distributed
- the populations have equal variances
- samples are independent

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-12

σ_1 and σ_2 unknown, small samples

(continued)

Population means, independent samples

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ *

Forming interval estimates:

- The population variances are assumed equal, so use the two sample standard deviations and pool them to estimate σ
- the test statistic is a t value with $(n_1 + n_2 - 2)$ degrees of freedom

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-13

σ_1 and σ_2 unknown, small samples

(continued)

Population means, independent samples

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ *

The pooled standard deviation is

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-14

σ_1 and σ_2 unknown, small samples

(continued)

Population means, independent samples

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ *

The confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where $t_{\alpha/2}$ has $(n_1 + n_2 - 2)$ d.f.,

and

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-15

Paired Samples

Paired samples

Tests Means of 2 Related Populations

- Paired or matched samples
- Repeated measures (before/after)
- Use difference between paired values:

$$d = x_1 - x_2$$

- Eliminates Variation Among Subjects
- Assumptions:
  - Both Populations Are Normally Distributed
  - Or, if Not Normal, use large samples

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-16

Paired Differences

Paired samples

The i[th] paired difference is d_i, where

$$d_i = x_{1i} - x_{2i}$$

The point estimate for the population mean paired difference is $\bar{d}$:

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

The sample standard deviation is

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

n is the number of pairs in the paired sample

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-17

Paired Differences

(continued)

Paired samples

The confidence interval for $\bar{d}$ is

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

Where $t_{\alpha/2}$ has n - 1 d.f. and s_d is:

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

n is the number of pairs in the paired sample

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-18

Hypothesis Tests for the Difference Between Two Means

- Testing Hypotheses about $\mu_1 - \mu_2$
- Use the same situations discussed already:
  - Standard deviations known or unknown
  - Sample sizes ≥ 30 or not ≥ 30

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-19

Hypothesis Tests for Two Population Proportions

Two Population Means, Independent Samples

Lower tail test:	Upper tail test:	Two-tailed test:
$H_0: \mu_1 \geq \mu_2$ $H_A: \mu_1 < \mu_2$ i.e., $H_0: \mu_1 - \mu_2 \geq 0$ $H_A: \mu_1 - \mu_2 < 0$	$H_0: \mu_1 \leq \mu_2$ $H_A: \mu_1 > \mu_2$ i.e., $H_0: \mu_1 - \mu_2 \leq 0$ $H_A: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 = \mu_2$ $H_A: \mu_1 \neq \mu_2$ i.e., $H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-20

Hypothesis tests for $\mu_1 - \mu_2$

Population means, independent samples

- σ_1 and σ_2 known → Use a **z** test statistic
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$ → Use s to estimate unknown σ, approximate with a **z** test statistic
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ → Use s to estimate unknown σ, use a **t** test statistic and pooled standard deviation

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-21

σ_1 and σ_2 known

Population means, independent samples

- σ_1 and σ_2 known *
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

The test statistic for $\mu_1 - \mu_2$ is:

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

σ_1 and σ_2 unknown, large samples

Population means, independent samples

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$ *
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$

The test statistic for $\mu_1 - \mu_2$ is:

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

σ_1 and σ_2 unknown, small samples

Population means, independent samples

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown, n_1 and $n_2 \geq 30$
- σ_1 and σ_2 unknown, n_1 or $n_2 < 30$ *

The test statistic for $\mu_1 - \mu_2$ is:

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

Where $t_{\alpha/2}$ has $(n_1 + n_2 - 2)$ d.f.,

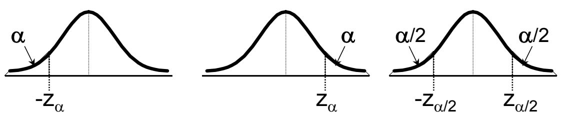
and

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

Hypothesis tests for $\mu_1 - \mu_2$

Two Population Means, Independent Samples

Lower tail test:	Upper tail test:	Two-tailed test:
$H_0: \mu_1 - \mu_2 \geq 0$ $H_A: \mu_1 - \mu_2 < 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_A: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$



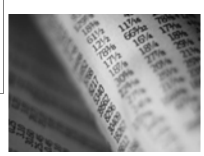
Reject H_0 if $z < -z_\alpha$	Reject H_0 if $z > z_\alpha$	Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-25

Pooled s_p t Test: Example

You're a financial analyst for a brokerage firm. Is there a difference in dividend yield between stocks listed on the NYSE & NASDAQ? You collect the following data:

	NYSE	NASDAQ
Number	**21**	**25**
Sample mean	**3.27**	**2.53**
Sample std dev	**1.30**	**1.16**



Assuming equal variances, is there a difference in average yield (α **= 0.05**)?

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-26

Calculating the Test Statistic

The test statistic is:

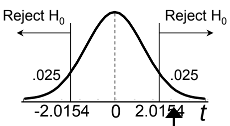
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(3.27 - 2.53) - 0}{1.2256 \sqrt{\frac{1}{21} + \frac{1}{25}}} = \boxed{2.040}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(21-1)1.30^2 + (25-1)1.16^2}{21 + 25 - 2}} = 1.2256$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-27

Solution

H_0: $\mu_1 - \mu_2 = 0$ i.e. ($\mu_1 = \mu_2$)
H_A: $\mu_1 - \mu_2 \neq 0$ i.e. ($\mu_1 \neq \mu_2$)
$\alpha = 0.05$
df = 21 + 25 - 2 = 44
Critical Values: t = ± 2.0154



Test Statistic:

$$z = \frac{3.27 - 2.53}{1.2256\sqrt{\frac{1}{21} + \frac{1}{25}}} = 2.040$$

Decision:
Reject H_0 at α = 0.05

Conclusion:
There is evidence of a difference in means.

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-28

Hypothesis Testing for Paired Samples

Paired samples

The test statistic for $\bar{d}$ is

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

n is the number of pairs in the paired sample

Where $t_{\alpha/2}$ has n - 1 d.f. and s_d is:

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

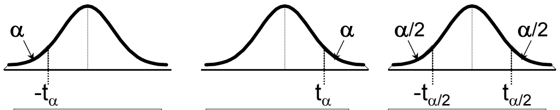
Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-29

Hypothesis Testing for Paired Samples

(continued)

Paired Samples

Lower tail test:	Upper tail test:	Two-tailed test:
H_0: $\mu_d \geq 0$ H_A: $\mu_d < 0$	H_0: $\mu_d \leq 0$ H_A: $\mu_d > 0$	H_0: $\mu_d = 0$ H_A: $\mu_d \neq 0$



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

Where t has n - 1 d.f.

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-30

Paired Samples Example

- Assume you send your salespeople to a "customer service" training workshop. Is the training effective? You collect the following data:

Salesperson	Number of Complaints: Before (1)	After (2)	(2) - (1) Difference, d_i
C.B.	6	4	- 2
T.F.	20	6	-14
M.H.	3	2	- 1
R.K.	0	0	0
M.O.	4	0	- 4
			-21

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

$= -4.2$

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

$= 5.67$

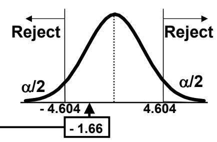
Paired Samples: Solution

- Has the training made a difference in the number of complaints (at the 0.01 level)?

H_0: $\mu_d = 0$
H_A: $\mu_d \neq 0$

$\alpha = .01$ $\bar{d} = -4.2$

Critical Value = ± 4.604
d.f. = n - 1 = 4



Test Statistic:

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-4.2 - 0}{5.67/\sqrt{5}} = -1.66$$

Decision: Do not reject H_0
(t stat is not in the reject region)

Conclusion: There is not a significant change in the number of complaints.

Two Population Proportions

Population proportions

Goal: Form a confidence interval for or test a hypothesis about the difference between two population proportions, $p_1 - p_2$

Assumptions:

$n_1 p_1 \geq 5$, $n_1(1-p_1) \geq 5$

$n_2 p_2 \geq 5$, $n_2(1-p_2) \geq 5$

The point estimate for the difference is $\bar{p}_1 - \bar{p}_2$

Confidence Interval for Two Population Proportions

Population proportions

The confidence interval for $p_1 - p_2$ is:

$$\left(\bar{p}_1 - \bar{p}_2\right) \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-34

Hypothesis Tests for Two Population Proportions

Population proportions

Lower tail test:	Upper tail test:	Two-tailed test:
$H_0: p_1 \geq p_2$ $H_A: p_1 < p_2$ i.e., $H_0: p_1 - p_2 \geq 0$ $H_A: p_1 - p_2 < 0$	$H_0: p_1 \leq p_2$ $H_A: p_1 > p_2$ i.e., $H_0: p_1 - p_2 \leq 0$ $H_A: p_1 - p_2 > 0$	$H_0: p_1 = p_2$ $H_A: p_1 \neq p_2$ i.e., $H_0: p_1 - p_2 = 0$ $H_A: p_1 - p_2 \neq 0$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-35

Two Population Proportions

Population proportions

Since we begin by assuming the null hypothesis is true, we assume $p_1 = p_2$ and pool the two $\bar{p}$ estimates

The pooled estimate for the overall proportion is:

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

where x_1 and x_2 are the numbers from samples 1 and 2 with the characteristic of interest

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-36

Two Population Proportions

(continued)

Population proportions

The test statistic for $p_1 - p_2$ is:

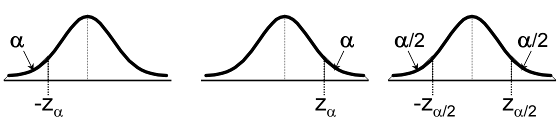
$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-37

Hypothesis Tests for Two Population Proportions

Population proportions

Lower tail test:	Upper tail test:	Two-tailed test:
$H_0: p_1 - p_2 \geq 0$ $H_A: p_1 - p_2 < 0$	$H_0: p_1 - p_2 \leq 0$ $H_A: p_1 - p_2 > 0$	$H_0: p_1 - p_2 = 0$ $H_A: p_1 - p_2 \neq 0$



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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-38

Example: Two population Proportions

Is there a significant difference between the proportion of men and the proportion of women who will vote Yes on Proposition A?

- In a random sample, 36 of 72 men and 31 of 50 women indicated they would vote Yes
- Test at the .05 level of significance

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 9-39

Example: Two population Proportions

(continued)

- The hypothesis test is:

H_0: $p_1 - p_2 = 0$ (the two proportions are equal)
H_A: $p_1 - p_2 \neq 0$ (there is a significant difference between proportions)

- The sample proportions are:
  - Men: $\bar{p}_1 = 36/72 = .50$
  - Women: $\bar{p}_2 = 31/50 = .62$
- The pooled estimate for the overall proportion is:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{36 + 31}{72 + 50} = \frac{67}{122} = .549$$

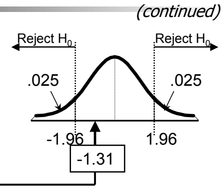
Example: Two population Proportions

(continued)

The test statistic for $p_1 - p_2$ is:

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

$$= \frac{(.50 - .62) - (0)}{\sqrt{.549\,(1 - .549)\left(\frac{1}{72} + \frac{1}{50}\right)}} = -1.31$$



Critical Values = ±1.96
For α = .05

Decision: Do not reject H_0

Conclusion: There is not significant evidence of a difference in proportions who will vote yes between men and women.

Two Sample Tests in EXCEL

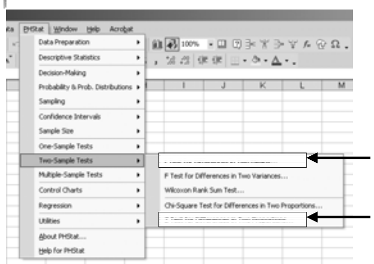
For independent samples:

- Independent sample Z test with variances known:
  - Tools | data analysis | z-test: two sample for means
- Independent sample Z test with large sample
  - Tools | data analysis | z-test: two sample for means
  - If the population variances are unknown, use sample variances

For paired samples (t test):

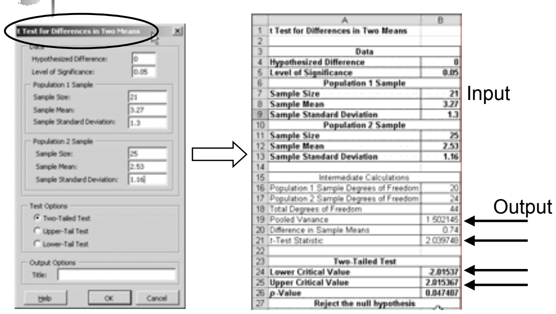
- Tools | data analysis… | t-test: paired two sample for means

Two Sample Tests in PHStat



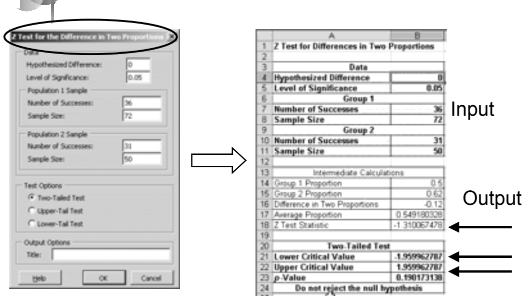
Chap 9-43

Two Sample Tests in PHStat



Chap 9-44

Two Sample Tests in PHStat



Chap 9-45

Chapter Summary

- Compared two independent samples
  - Formed confidence intervals for the differences between two means
  - Performed Z test for the differences in two means
  - Performed t test for the differences in two means
- Compared two related samples (paired samples)
  - Formed confidence intervals for the paired difference
  - Performed paired sample t tests for the mean difference
- Compared two population proportions
  - Formed confidence intervals for the difference between two population proportions
  - Performed Z-test for two population proportions