Statistics for Business and Economics

7th Edition

Chapter 10

Hypothesis Testing: Additional Topics

Chapter Goals

After completing this chapter, you should be able to:

- Test hypotheses for the difference between two population means
  - Two means, matched pairs
  - Independent populations, population variances known
  - Independent populations, population variances unknown but equal
- Complete a hypothesis test for the difference between two proportions (large samples)
- Use the chi-square distribution for tests of the variance of a normal distribution
- Use the F table to find critical F values
- Complete an F test for the equality of two variances

Two Sample Tests

Two Sample Tests

- Population Means, Dependent Samples
- Population Means, Independent Samples
- Population Proportions
- Population Variances

Examples:

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

10.1

Dependent Samples

Dependent Samples

Tests Means of 2 Related Populations

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

$$d_i = x_i - y_i$$

- Assumptions:
  - Both Populations Are Normally Distributed

Test Statistic: Dependent Samples

Dependent Samples

The test statistic for the mean difference is a t value, with n – 1 degrees of freedom:

$$t = \frac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}}$$

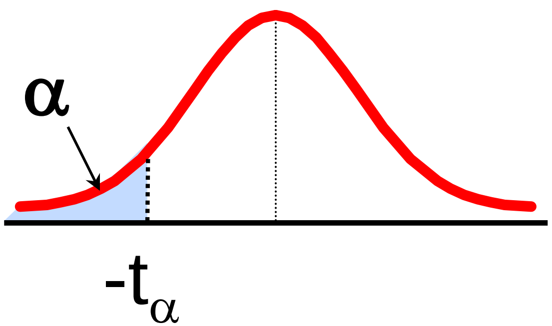
where $\bar{d} = \frac{\sum d_i}{n} = \bar{x} - \bar{y}$

D_0 = hypothesized mean difference
s_d = sample standard dev. of differences
n = the sample size (number of pairs)

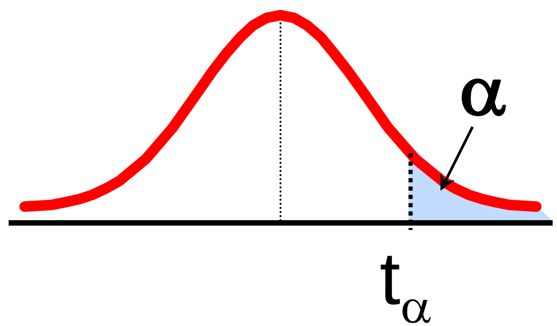
Decision Rules: Matched Pairs

Matched or Paired Samples

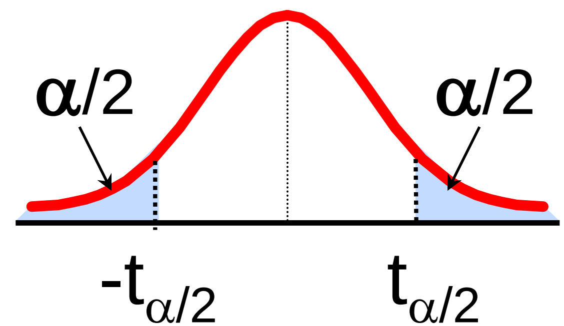
Lower-tail test:	Upper-tail test:	Two-tail test:
H_0: $\mu_x - \mu_y \geq 0$ H_1: $\mu_x - \mu_y < 0$	H_0: $\mu_x - \mu_y \leq 0$ H_1: $\mu_x - \mu_y > 0$	H_0: $\mu_x - \mu_y = 0$ H_1: $\mu_x - \mu_y \neq 0$



Reject H_0 if $t < -t_{n-1,\, \alpha}$



Reject H_0 if $t > t_{n-1,\, \alpha}$



Reject H_0 if $t < -t_{n-1\,,\, \alpha/2}$
or $t > t_{n-1\,,\, \alpha/2}$

Where $t = \dfrac{\bar{d} - D_0}{\dfrac{s_d}{\sqrt{n}}}$ has n - 1 d.f.

Matched Pairs Example

- Assume you send your salespeople to a "customer service" training workshop. Has the training made a difference in the number of complaints? 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

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

$$= -4.2$$

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

$$= 5.67$$

Matched Pairs: Solution

- Has the training made a difference in the number of complaints (at the α = 0.05 level)?

$$H_0: \mu_x - \mu_y = 0$$
$$H_1: \mu_x - \mu_y \neq 0$$

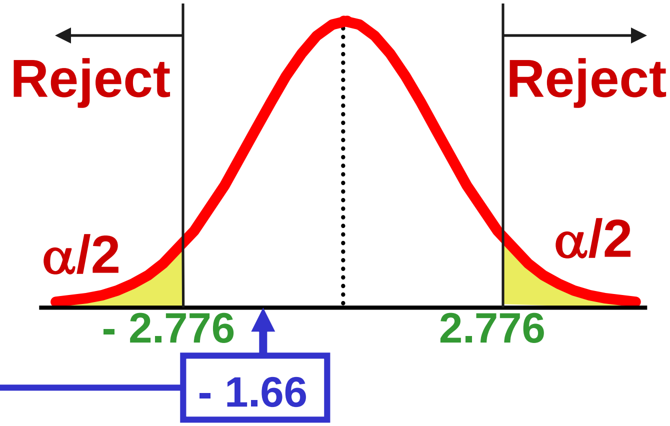
$\alpha = .05 \qquad \bar{d} = -4.2$

Critical Value = ± 2.776

d.f. = n − 1 = 4

Test Statistic:

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}} = \frac{-4.2 - 0}{5.67/\sqrt{5}} = \boxed{-1.66}$$



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.

10.2

Difference Between Two Means

Population means, independent samples

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

- Different populations
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
  - Normally distributed

Difference Between Two Means

(continued)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known → Test statistic is a z value
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal → Test statistic is a a value from the Student's t distribution
    - σ_x^2 and σ_y^2 assumed unequal → Test statistic is a a value from the Student's t distribution

σ_x^2 and σ_y^2 Known

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Assumptions:

- Samples are randomly and independently drawn
- both population distributions are normal
- Population variances are known

σ_x^2 and σ_y^2 Known

(continued)

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

When σ_x^2 and σ_y^2 are known and both populations are normal, the variance of $\overline{X} - \overline{Y}$ is

$$\sigma^2_{\overline{X}-\overline{Y}} = \frac{\sigma_x^{\ 2}}{n_x} + \frac{\sigma_y^{\ 2}}{n_y}$$

…and the random variable

$$Z = \frac{(\overline{x} - \overline{y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_x^2}{n_X} + \dfrac{\sigma_y^2}{n_Y}}}$$

has a standard normal distribution

Test Statistic, σ_x^2 and σ_y^2 Known

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

$$H_0 : \mu_x - \mu_y = D_0$$

The test statistic for $\mu_x - \mu_y$ is:

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$$

Hypothesis Tests for Two Population Means

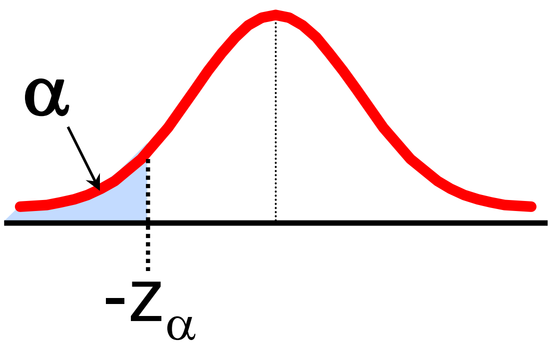
Two Population Means, Independent Samples

Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: \mu_x \geq \mu_y$ $H_1: \mu_x < \mu_y$	$H_0: \mu_x \leq \mu_y$ $H_1: \mu_x > \mu_y$	$H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$
i.e.,	i.e.,	i.e.,
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$

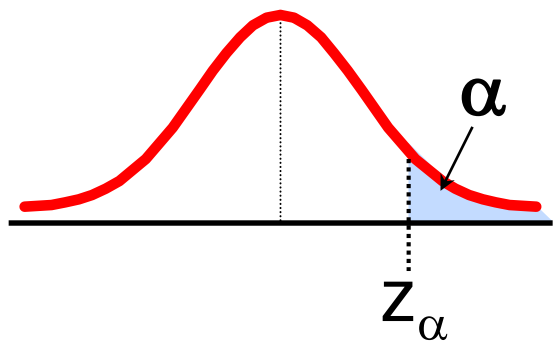
Decision Rules

Two Population Means, Independent Samples, Variances Known

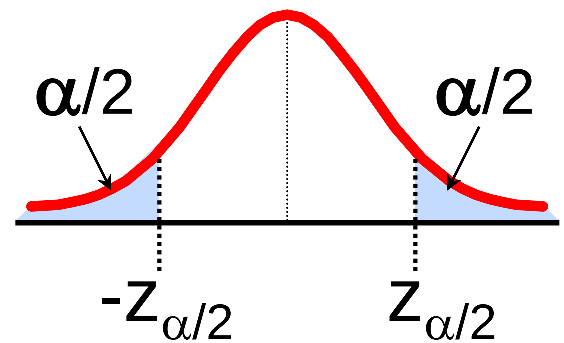
Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$



Reject H_0 if $z < -z_\alpha$



Reject H_0 if $z > z_\alpha$



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

σ_x^2 and σ_y^2 Unknown, Assumed Equal

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown but assumed equal

σ_x^2 and σ_y^2 Unknown, Assumed Equal

(continued)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

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

Test Statistic, σ_x^2 and σ_y^2 Unknown, Equal

σ_x^2 and σ_y^2 unknown

- σ_x^2 and σ_y^2 assumed equal *
- σ_x^2 and σ_y^2 assumed unequal

The test statistic for $\mu_x - \mu_y$ is:

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_p^2}{n_x} + \dfrac{s_p^2}{n_y}}}$$

Where t has $(n_1 + n_2 - 2)$ d.f.,

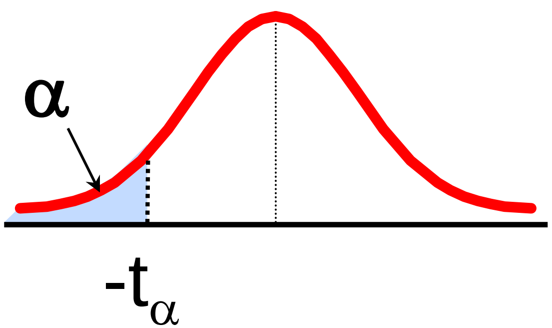
and

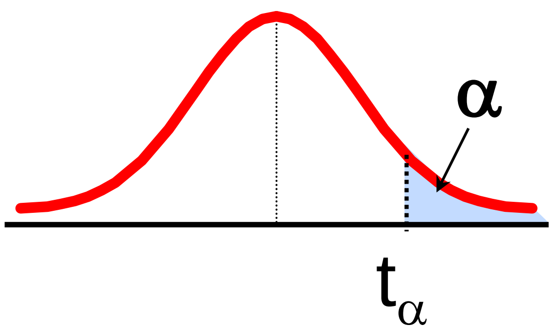
$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

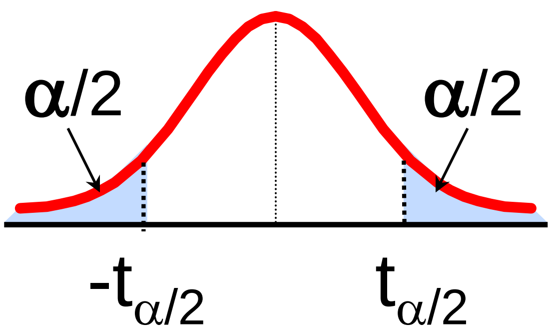
Decision Rules

Two Population Means, Independent Samples, Variances Unknown

Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$







Reject H_0 if
$t < -t_{(n1+n2-2), \alpha}$

Reject H_0 if
$t > t_{(n1+n2-2), \alpha}$

Reject H_0 if
$t < -t_{(n1+n2-2), \alpha/2}$ or
$t > t_{(n1+n2-2), \alpha/2}$

Pooled Variance t Test: Example

You are 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 both populations are approximately normal with equal variances, is there a difference in average yield (α = 0.05)?

Calculating the Test Statistic

The test statistic is:

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\left(3.27 - 2.53\right) - 0}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

$$S_p^2 = \frac{\left(n_1 - 1\right)S_1^{\,2} + \left(n_2 - 1\right)S_2^{\,2}}{\left(n_1 - 1\right) + \left(n_2 - 1\right)} = \frac{\left(21 - 1\right)1.30^2 + \left(25 - 1\right)1.16^2}{\left(21 - 1\right) + \left(25 - 1\right)} = 1.5021$$

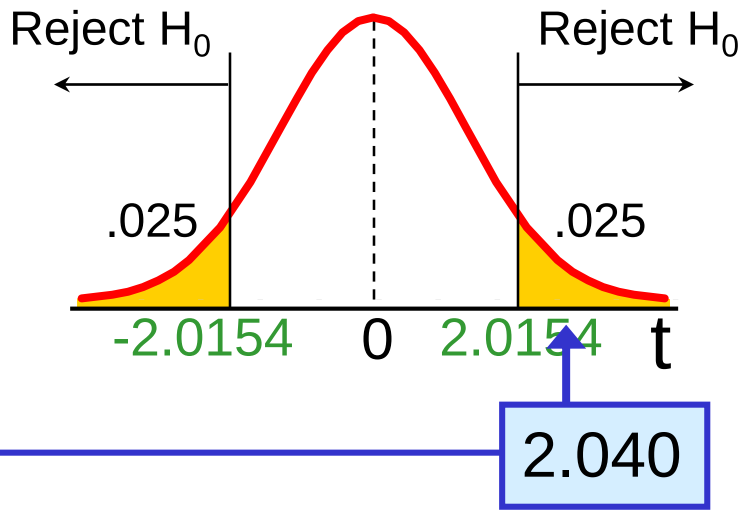
Solution

H_0: $\mu_1 - \mu_2 = 0$ i.e. ($\mu_1 = \mu_2$)
H_1: $\mu_1 - \mu_2 \neq 0$ i.e. ($\mu_1 \neq \mu_2$)

$\alpha = 0.05$

df = 21 + 25 − 2 = 44

Critical Values: t = ± 2.0154



Test Statistic:

$$t = \frac{3.27 - 2.53}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

There is evidence of a difference in means.

σ_x^2 and σ_y^2 Unknown, Assumed Unequal

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal
    - σ_x^2 and σ_y^2 assumed unequal *

Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown and assumed unequal

σ_x^2 and σ_y^2 Unknown, Assumed Unequal

(continued)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal
    - σ_x^2 and σ_y^2 assumed unequal *

Forming interval estimates:

- The population variances are assumed unequal, so a pooled variance is not appropriate
- use a t value with ν degrees of freedom, where

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right)+\left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2/(n_x-1)+\left(\frac{s_y^2}{n_y}\right)^2/(n_y-1)}$$

Test Statistic, σ_x^2 and σ_y^2 Unknown, Unequal

σ_x^2 and σ_y^2 unknown

- σ_x^2 and σ_y^2 assumed equal
- σ_x^2 and σ_y^2 assumed unequal *

The test statistic for $\mu_x - \mu_y$ is:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\dfrac{s_x^2}{n_X} + \dfrac{s_y^2}{n_Y}}}$$

Where t has ν degrees of freedom:

$$\nu = \frac{\left[\left(\dfrac{s_x^2}{n_x}\right) + \left(\dfrac{s_y^2}{n_y}\right)\right]^2}{\left(\dfrac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\dfrac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$