Statistics for Business and Economics

8th 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 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:

Population Means, Dependent Samples	Population Means, Independent Samples	Population Proportions	Population Variances
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 of the Difference Between Two Normal Population Means: 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

Population Means, Dependent Samples

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

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

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

s_d = sample standard dev. of differences
n = the sample size (number of pairs)

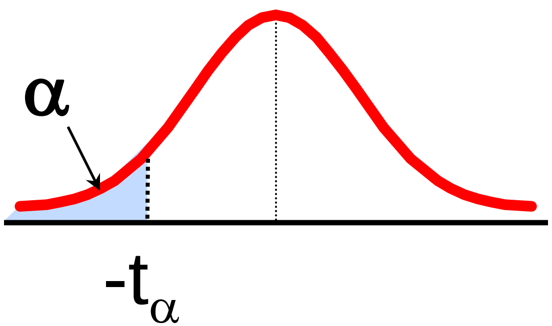
For tests of the following form:

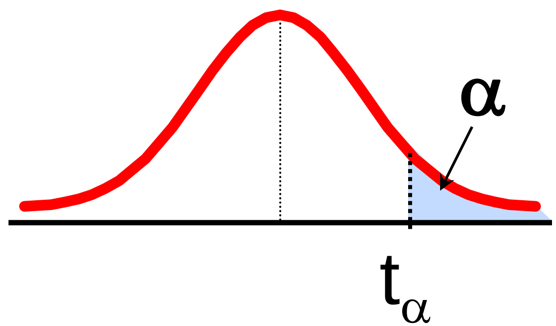
H_0: $\mu_x - \mu_y \geq 0$
H_0: $\mu_x - \mu_y \leq 0$
H_0: $\mu_x - \mu_y = 0$

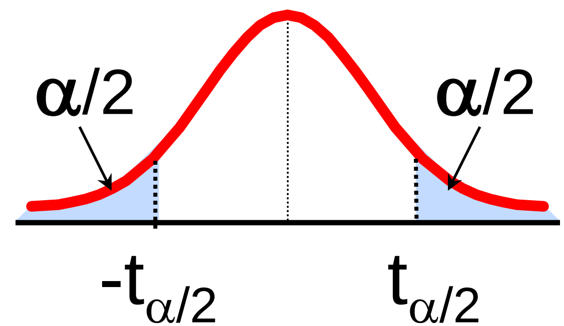
Decision Rules: Matched Pairs

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 $t = \dfrac{\bar{d}}{\dfrac{s_d}{\sqrt{n}}}$ 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

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

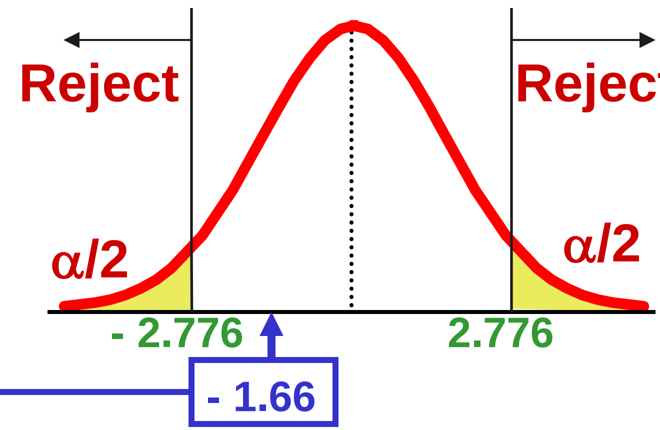
$$= -4.2$$

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

$$= 5.67$$

Matched Pairs: Solution

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



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

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

Critical Value = ± 2.776

d.f. = n − 1 = 4

Test Statistic:

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

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.

10.2

Independent Samples

Tests of the Difference Between Two Normal Population Means: Dependent Samples

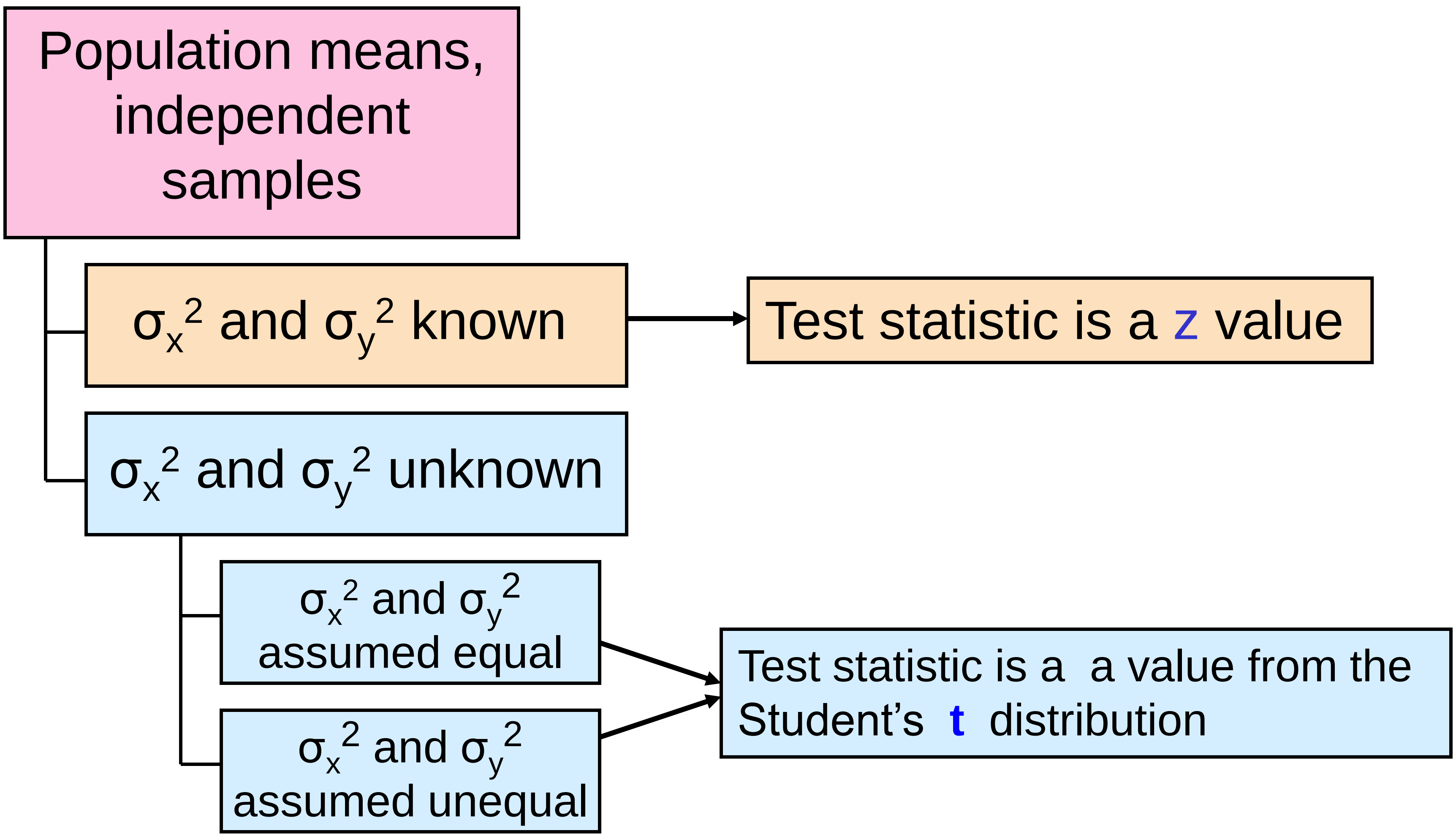
Population means, independent samples

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)



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 $\bar{X} - \bar{Y}$ is

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

…and the random variable

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

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

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

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

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

Hypothesis Tests for Two Population Means

Two Population Means, Independent Samples

Lower-tail test:

$H_0: \mu_x \geq \mu_y$
$H_1: \mu_x < \mu_y$

i.e.,

$H_0: \mu_x - \mu_y \geq 0$
$H_1: \mu_x - \mu_y < 0$

Upper-tail test:

$H_0: \mu_x \leq \mu_y$
$H_1: \mu_x > \mu_y$

i.e.,

$H_0: \mu_x - \mu_y \leq 0$
$H_1: \mu_x - \mu_y > 0$

Two-tail test:

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

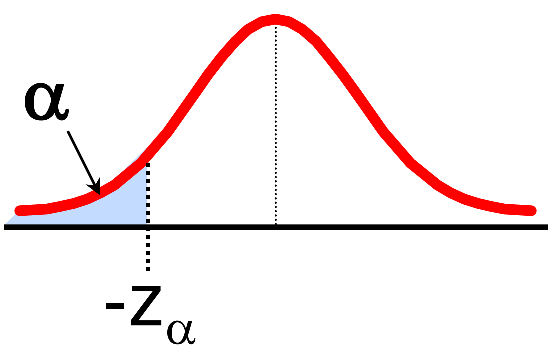
i.e.,

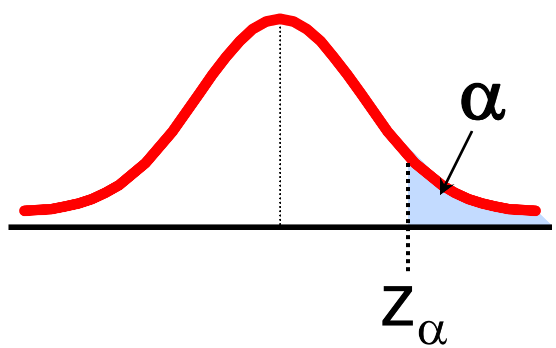
$H_0: \mu_x - \mu_y = 0$
$H_1: \mu_x - \mu_y \neq 0$

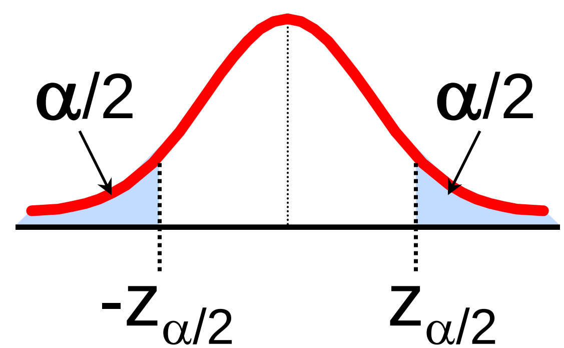
Decision Rules

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 $H_0: \mu_x - \mu_y = 0$ is:

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

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

and

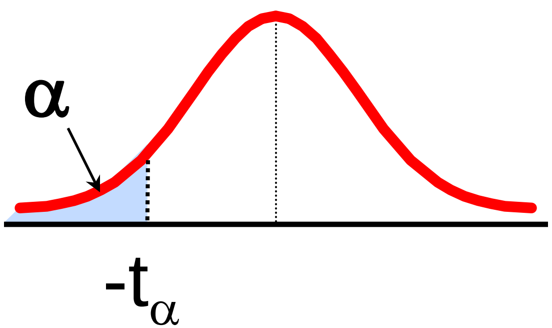
$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Decision Rules

Two Population Means, Independent Samples, Variances Unknown

Lower-tail test:

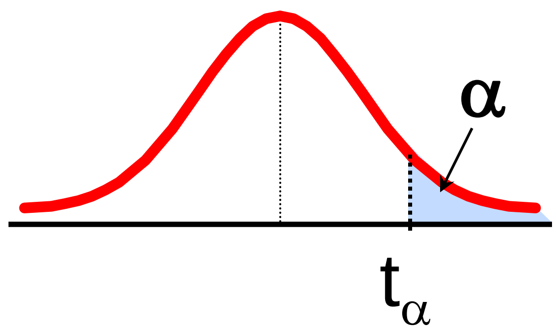
$H_0: \mu_x - \mu_y \geq 0$
$H_1: \mu_x - \mu_y < 0$



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

Upper-tail test:

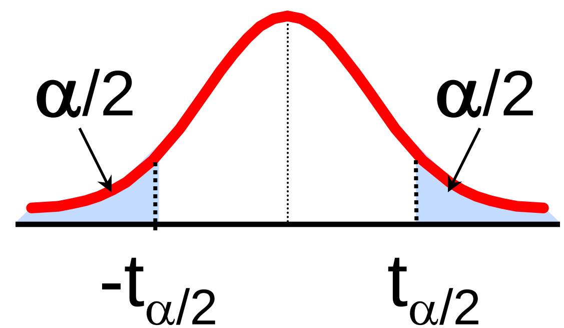
$H_0: \mu_x - \mu_y \leq 0$
$H_1: \mu_x - \mu_y > 0$



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

Two-tail test:

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



Reject H_0 if
$t < -t_{(n1+n2-2), \alpha/2}$ 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 ($\alpha = 0.05$)?

Calculating the Test Statistic

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

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

The test statistic is:

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

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

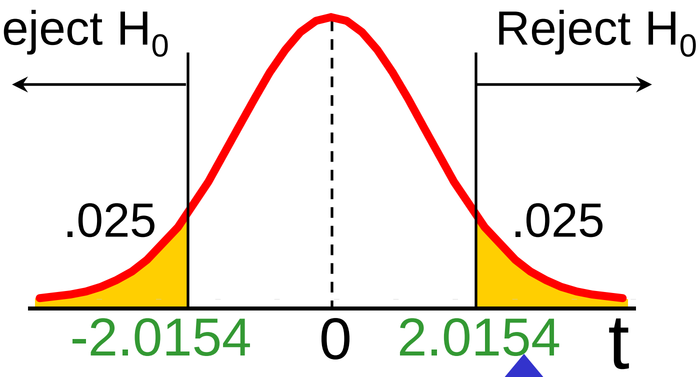
Solution

$H_0: \mu_1 - \mu_2 = 0$ 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: $t = \pm 2.0154$



Test Statistic:

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

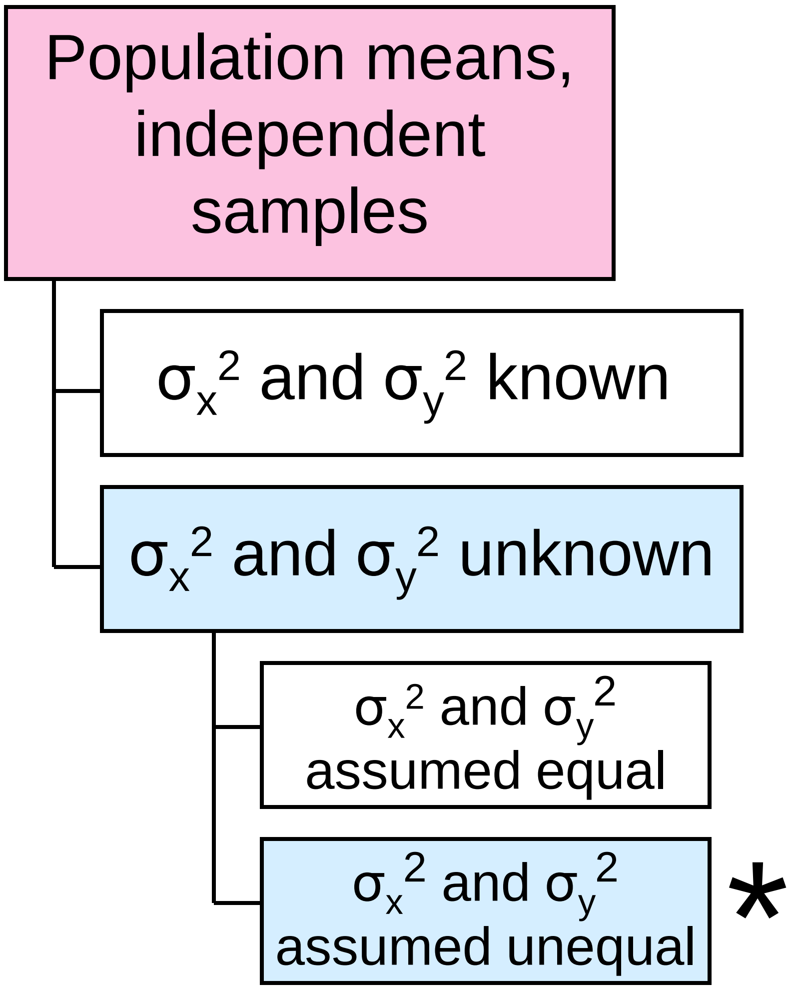
Decision:

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



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

H_0: $\mu_x - \mu_y = 0$ is:

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

Where t has ν degrees of freedom:

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

10.3

Two Population Proportions

Tests of the Difference Between Two Population Proportions (Large Samples)

Population proportions

Goal: Test hypotheses for the difference between two population proportions, $P_x - P_y$

Assumptions:

Both sample sizes are large,

$$nP(1 - P) > 5$$

Two Population Proportions

(continued)

Population proportions

- The random variable

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (P_x - P_y)}{\sqrt{\dfrac{P_x(1-P_x)}{n_x} + \dfrac{P_y(1-P_y)}{n_y}}}$$

has a standard normal distribution

Test Statistic for Two Population Proportions

Population proportions

The test statistic for

$H_0: P_x - P_y = 0$

is a z value:

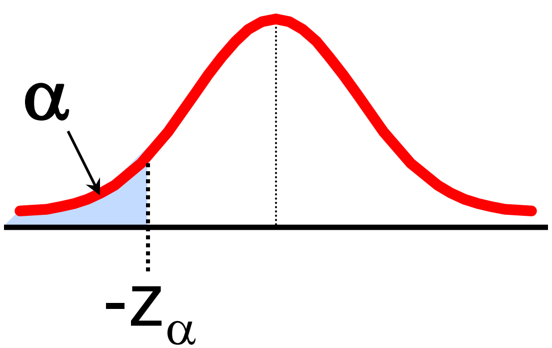
$$z = \frac{\left(\hat{p}_x - \hat{p}_y\right)}{\sqrt{\dfrac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \dfrac{\hat{p}_0(1-\hat{p}_0)}{n_y}}}$$

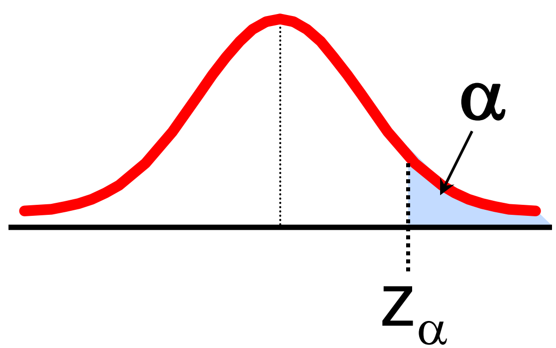
Where $$\hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y}$$

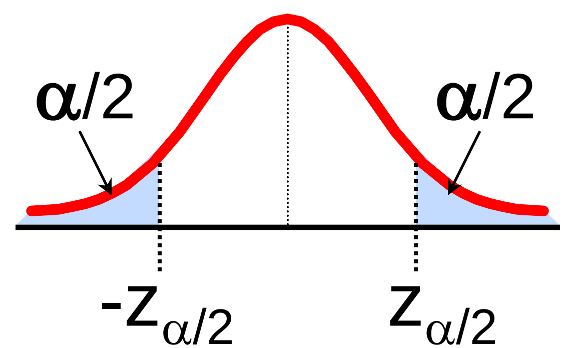
Decision Rules: Proportions

Population proportions

Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: P_x – P_y \geq 0$ $H_1: P_x – P_y < 0$	$H_0: P_x – P_y \leq 0$ $H_1: P_x – P_y > 0$	$H_0: P_x – P_y = 0$ $H_1: P_x – P_y \neq 0$







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

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

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

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

Example: Two Population Proportions

(continued)

- The hypothesis test is:

$H_0: P_M - P_W = 0$ (the two proportions are equal)

$H_1: P_M - P_W \neq 0$ (there is a significant difference between proportions)

- The sample proportions are:
  - Men: $\hat{p}_M = 36/72 = .50$
  - Women: $\hat{p}_W = 31/50 = .62$

- The estimate for the common overall proportion is:

$$\hat{p}_0 = \frac{n_M \hat{p}_M + n_W \hat{p}_W}{n_M + n_W} = \frac{72(36/72) + 50(31/50)}{72 + 50} = \frac{67}{122} = .549$$

Example: Two Population Proportions

(continued)

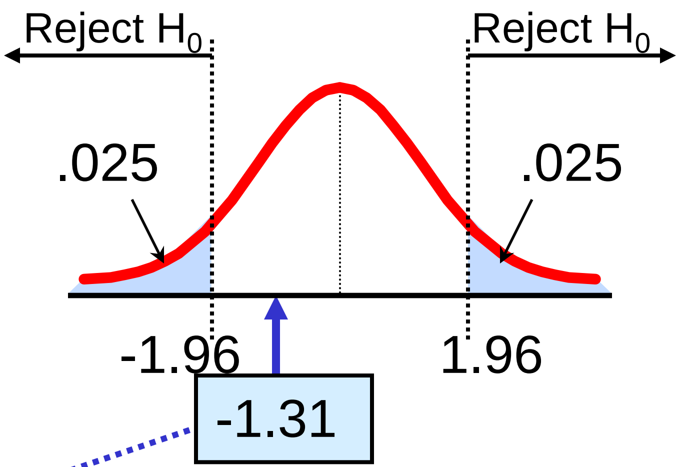
The test statistic for $P_M - P_W = 0$ is:

$$z = \frac{\left(\hat{p}_M - \hat{p}_W\right)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_1} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_2}}}$$

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

$$= -1.31$$

Critical Values = ±1.96
For α = .05



Decision: Do not reject H_0

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

10.4

Tests of Equality of Two Variances

Tests for Two Population Variances

F test statistic

- Goal: Test hypotheses about two population variances

$H_0: \sigma_x^2 \geq \sigma_y^2$
$H_1: \sigma_x^2 < \sigma_y^2$

Lower-tail test

$H_0: \sigma_x^2 \leq \sigma_y^2$
$H_1: \sigma_x^2 > \sigma_y^2$

Upper-tail test

$H_0: \sigma_x^2 = \sigma_y^2$
$H_1: \sigma_x^2 \neq \sigma_y^2$

Two-tail test

The two populations are assumed to be independent and normally distributed

Hypothesis Tests for Two Variances

(continued)

Tests for Two Population Variances

F test statistic

The random variable

$$F = \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2}$$

Has an F distribution with $(n_x - 1)$ numerator degrees of freedom and $(n_y - 1)$ denominator degrees of freedom

Denote an F value with ν_1 numerator and ν_2 denominator degrees of freedom by F_{ν_1,ν_2}

Test Statistic

Tests for Two Population Variances

F test statistic

The critical value for a hypothesis test about two population variances is

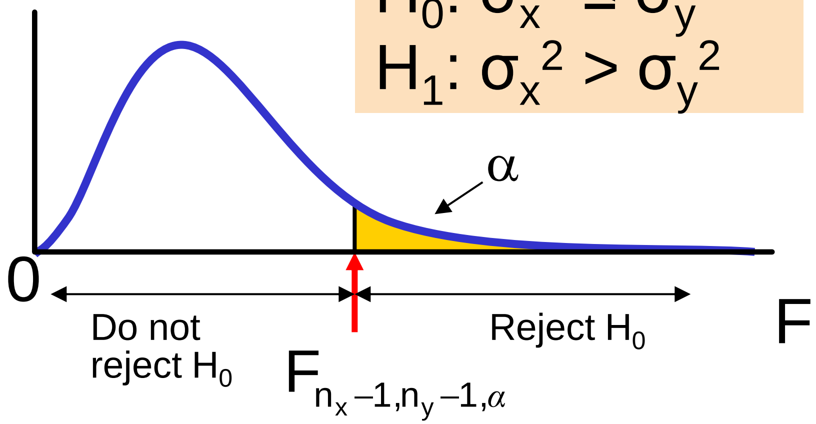
$$F = \frac{s_x^2}{s_y^2}$$

where F has $(n_x - 1)$ numerator degrees of freedom and $(n_y - 1)$ denominator degrees of freedom

Decision Rules: Two Variances

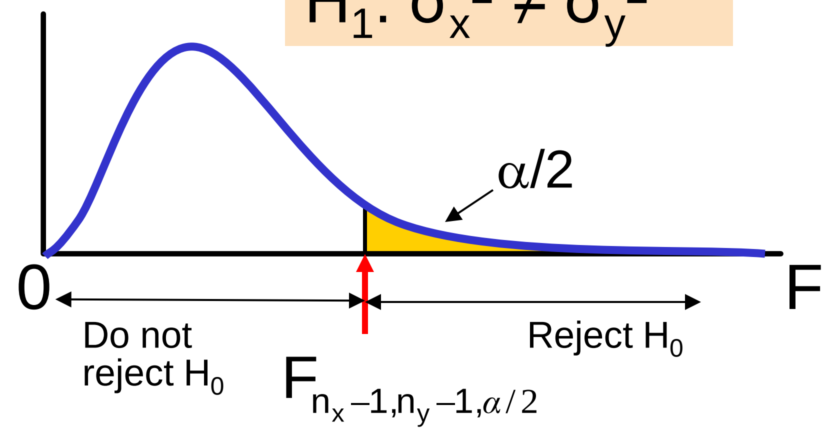
Use s_x^2 to denote the larger variance.

$H_0: \sigma_x^2 \le \sigma_y^2$
$H_1: \sigma_x^2 > \sigma_y^2$



$$\text{Reject } H_0 \text{ if } F > F_{n_x-1, n_y-1, \alpha}$$

$H_0: \sigma_x^2 = \sigma_y^2$
$H_1: \sigma_x^2 \ne \sigma_y^2$



- rejection region for a two-tail test is:

$$\text{Reject } H_0 \text{ if } F > F_{n_x-1, n_y-1, \alpha/2}$$

where s_x^2 is the larger of the two sample variances

Example: F Test

You are a financial analyst for a brokerage firm. You want to compare dividend yields between stocks listed on the NYSE & NASDAQ. You collect the following data:

	NYSE	NASDAQ
Number	**21**	**25**
Mean	**3.27**	**2.53**
Std dev	**1.30**	**1.16**

Is there a difference in the variances between the NYSE & NASDAQ at the $\alpha = 0.10$ level?

F Test: Example Solution

- Form the hypothesis test:

$H_0: \sigma_x^2 = \sigma_y^2$ (there is no difference between variances)

$H_1: \sigma_x^2 \neq \sigma_y^2$ (there is a difference between variances)

- Find the F critical values for $\alpha = .10/2$:

Degrees of Freedom:

- Numerator

  (NYSE has the larger standard deviation):

  - $n_x - 1 = 21 - 1 = 20$ d.f.

- Denominator:

  - $n_y - 1 = 25 - 1 = 24$ d.f.

$$F_{n_x-1,\, n_y-1,\, \alpha/2}$$

$$= F_{20,\, 24,\, 0.10/2} = 2.03$$

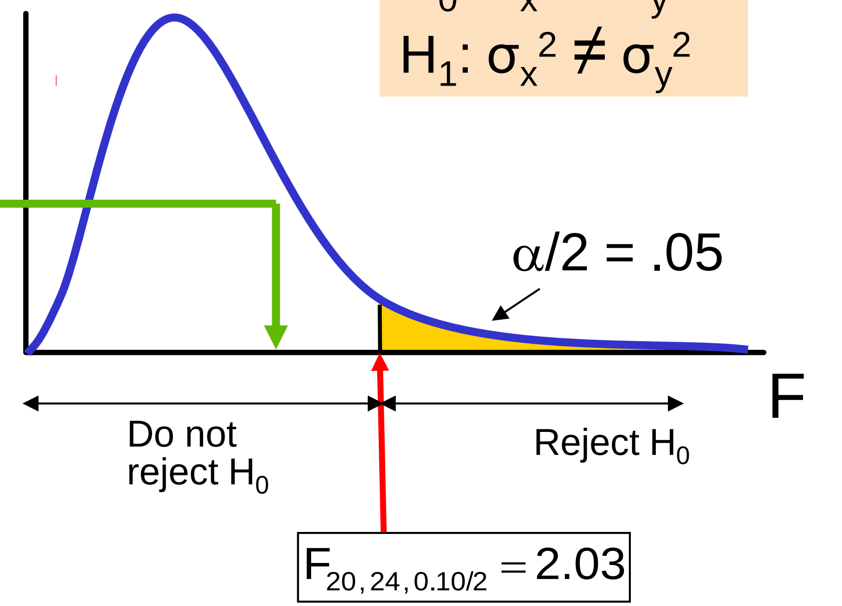
F Test: Example Solution

(continued)

- The test statistic is:

$$H_0: \sigma_x^2 = \sigma_y^2$$
$$H_1: \sigma_x^2 \neq \sigma_y^2$$

$$F = \frac{s_x^2}{s_y^2} = \frac{1.30^2}{1.16^2} = 1.256$$



- F = 1.256 is not in the rejection region, so we do not reject H_0

- Conclusion: There is not sufficient evidence of a difference in variances at α = .10

Some Comments on Hypothesis Testing

- A test with low power can result from:
  - Small sample size
  - Large variances in the underlying populations
  - Poor measurement procedures
- If sample sizes are large it is possible to find significant differences that are not practically important
- Researchers should select the appropriate level of significance before computing p-values

Two-Sample Tests in EXCEL

For paired samples (t test):

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

For independent samples:

- Independent sample z test with variances known:
  - Data | data analysis | z-test: two sample for means

For variances…

- F test for two variances:
  - Data | data analysis | F-test: two sample for variances

Chapter Summary

- Compared two dependent samples (paired samples)
  - Performed paired sample t test for the mean difference
- Compared two independent samples
  - Performed z test for the differences in two means
  - Performed pooled variance t test for the differences in two means
- Compared two population proportions
  - Performed z-test for two population proportions

Chapter Summary

(continued)

- Performed F tests for the difference between two population variances
- Used the F table to find F critical values