

# Statistik Bisnis

Week 11

Two-Sample Tests

# Learning Objectives

The means of two independent populations

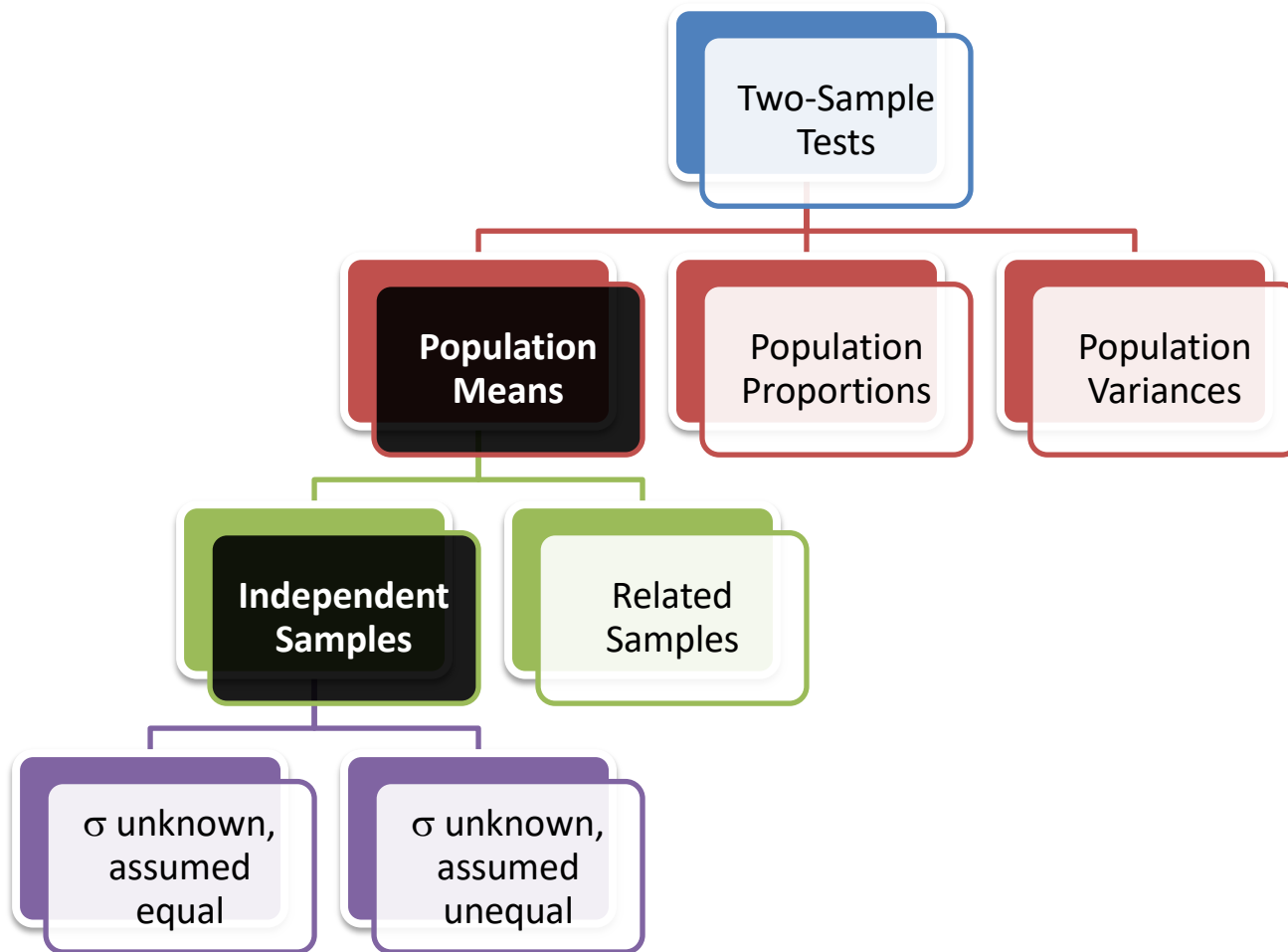
The means of two related populations

**In this chapter, you learn how to use hypothesis testing for comparing the difference between:**

The proportions of two independent populations

The variances of two independent populations by testing the ratio of the two variances

# Two-Sample Tests



# Difference Between Two Means

## Independent Samples

$\sigma$  unknown,  
assumed  
equal

$\sigma$  unknown,  
assumed  
unequal

**Goal:** Test hypothesis or form a confidence interval for the difference between two population means,  $\mu_1 - \mu_2$

The point estimate for the difference is

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

# Difference Between Two Means: Independent Samples

Different data sources

- Unrelated
- Independent
  - Sample selected from one population has no effect on the sample selected from the other population

Independent  
Samples

$\sigma$  unknown,  
assumed  
equal

Use  $S_p$  to estimate unknown  $\sigma$ . Use a **Pooled-Variance t test.**

$\sigma$  unknown,  
assumed  
unequal

Use  $S_1$  and  $S_2$  to estimate unknown  $\sigma_1$  and  $\sigma_2$ . Use a **Separate-Variance t test.**

# Hypothesis Tests for Two Population Means

## Two Population Means, Independent Samples

Lower-tail test:

$$H_0: \mu_1 \geq \mu_2$$

$$H_1: \mu_1 < \mu_2$$

i.e.,

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

$$H_1: \mu_1 - \mu_2 < 0$$

Upper-tail test:

$$H_0: \mu_1 \leq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

i.e.,

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

Two-tail test:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

i.e.,

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

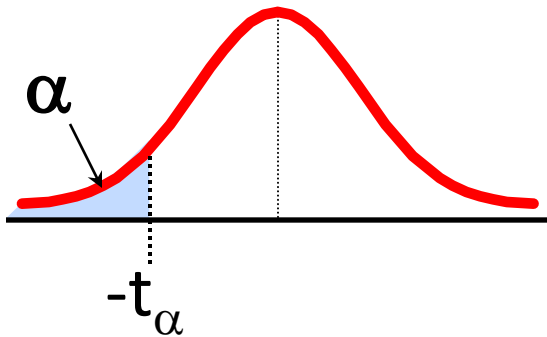
# Hypothesis tests for $\mu_1 - \mu_2$

## Two Population Means, Independent Samples

Lower-tail test:

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

$$H_1: \mu_1 - \mu_2 < 0$$

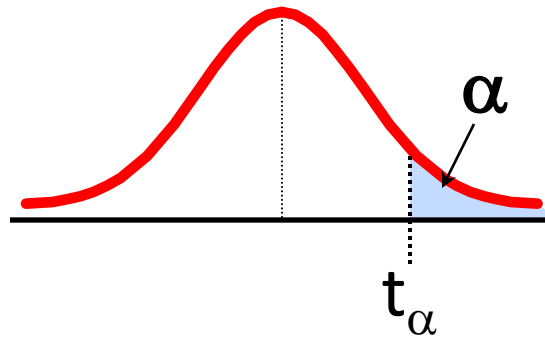


Reject  $H_0$  if  $t_{\text{STAT}} < -t_\alpha$

Upper-tail test:

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

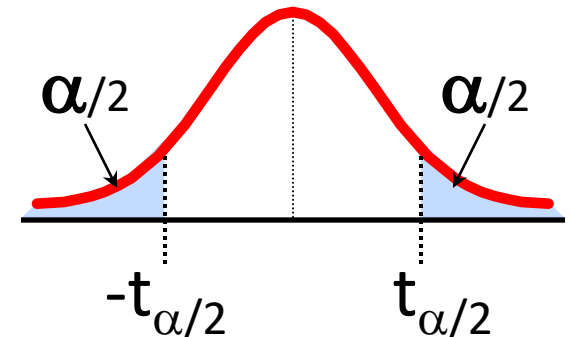


Reject  $H_0$  if  $t_{\text{STAT}} > t_\alpha$

Two-tail test:

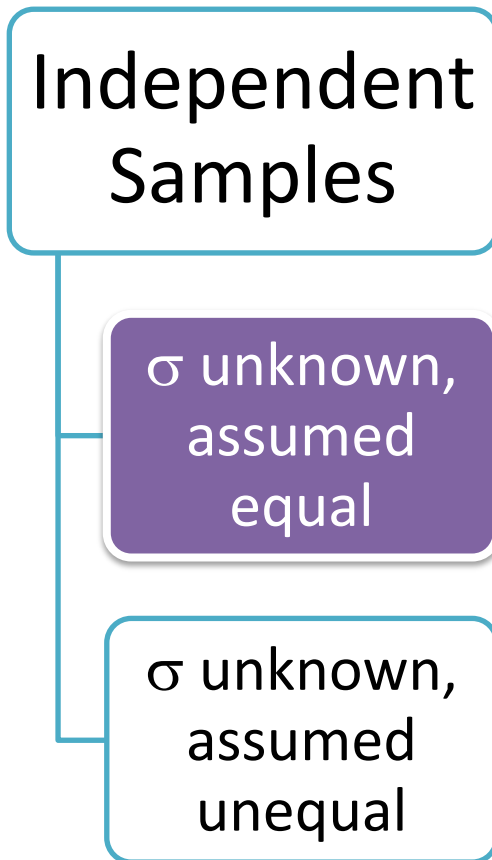
$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$



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

# Hypothesis tests for $\mu_1 - \mu_2$ with $\sigma_1$ and $\sigma_2$ unknown and assumed equal



## Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed or both sample sizes are at least 30
- Population variances are unknown but assumed equal



# Hypothesis tests for $\mu_1 - \mu_2$ with $\sigma_1$ and $\sigma_2$ unknown and assumed equal

*(continued)*

Independent  
Samples

$\sigma$  unknown,  
assumed  
equal

$\sigma$  unknown,  
assumed  
unequal

- The pooled variance is:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

- The test statistic is:

$$t_{\text{STAT}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- Where  $t_{\text{STAT}}$  has **d.f. =  $(n_1 + n_2 - 2)$**

# Confidence interval for $\mu_1 - \mu_2$ with $\sigma_1$ and $\sigma_2$ unknown and assumed equal

Independent  
Samples

$\sigma$  unknown,  
assumed  
equal

$\sigma$  unknown,  
assumed  
unequal

The confidence interval for

$\mu_1 - \mu_2$  is:

$$\left( \bar{X}_1 - \bar{X}_2 \right) \pm t_{\alpha/2} \sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Where  $t_{\alpha/2}$  has **d.f. =  $n_1 + n_2 - 2$**

# Pooled-Variance t Test Example

Anda adalah seorang analis keuangan dan ingin mengetahui apakah terdapat perbedaan hasil dividen antara saham yang terdaftar di NYSE & NASDAQ? Anda mengumpulkan data:

	<u>NYSE</u>	<u>NASDAQ</u>
<b>Jumlah</b>	<b>21</b>	<b>25</b>
<b>Rata-rata sampel</b>	<b>3.27</b>	<b>2.53</b>
<b>Std dev sampel</b>	<b>1.30</b>	<b>1.16</b>



Dengan mengasumsikan bahwa kedua populasi berdistribusi normal dengan variansi sama, apakah terdapat perbedaan pada rata-rata hasil ( $\alpha = 0.05$ )?

# Pooled-Variance t Test Example: Calculating the Test Statistic

*(continued)*

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

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

The test statistic is:

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

---

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

# Pooled-Variance t Test Example: Hypothesis Test Solution

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

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

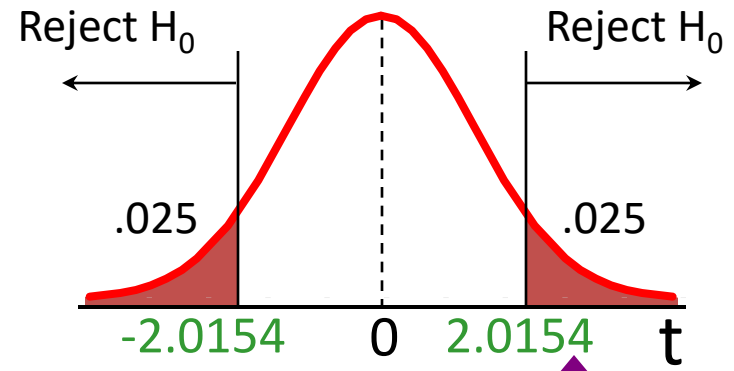
$$\alpha = 0.05$$

$$df = 21 + 25 - 2 = 44$$

$$\text{Critical Values: } t = \pm 2.0154$$

**Test Statistic:**

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



**Decision:**

Reject  $H_0$  at  $\alpha = 0.05$

**Conclusion:**

There is evidence of a difference in means.

# Pooled-Variance t Test Example: Confidence Interval for $\mu_1 - \mu_2$

DCOVA A

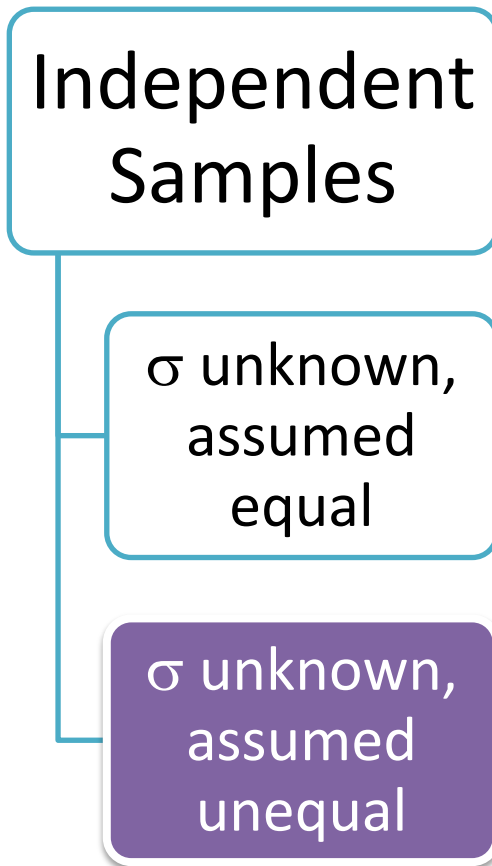
Since we rejected  $H_0$  can we be 95% confident that  $\mu_{\text{NYSE}} > \mu_{\text{NASDAQ}}$ ?

95% Confidence Interval for  $\mu_{\text{NYSE}} - \mu_{\text{NASDAQ}}$

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.74 \pm 2.0154 \times 0.3628 = (0.009, 1.471)$$

Since 0 is less than the entire interval, we can be 95% confident that  $\mu_{\text{NYSE}} > \mu_{\text{NASDAQ}}$

# Hypothesis tests for $\mu_1 - \mu_2$ with $\sigma_1$ and $\sigma_2$ unknown, not assumed equal



## Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed or both sample sizes are at least 30
- Population variances are unknown and cannot be assumed to be equal

# Hypothesis tests for $\mu_1 - \mu_2$ with $\sigma_1$ and $\sigma_2$ unknown and not assumed equal *(continued)*

Independent Samples

$\sigma$  unknown, assumed equal

$\sigma$  unknown, assumed unequal

The test statistic is:

$$t_{\text{STAT}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$t_{\text{STAT}}$  has d.f.  $\nu =$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$



**EXERCISE**

## 10.8 (cont'd)

Menurut penelitian terbaru, anak-anak yang menonton kartun dengan iklan sebuah produk makanan, rata-rata, makan 28,5 gram Goldfish crackers. Sementara itu, anak-anak yang menonton kartun tanpa iklan tersebut rata-rata makan 19,7 gram Goldfish crackers. Terdapat 118 anak dalam penelitian ini, yaitu 59 anak untuk setiap kelompok. Simpangan baku sampel untuk anak yang menonton kartun dengan iklan adalah 8,6 gram dan simpangan baku sampel untuk anak menonton tanpa iklan adalah 7,9 gram.

## 10.8

Dengan mengasumsikan variansi populasi sama dan  $\alpha=0,05$ , apakah terdapat cukup bukti bahwa rata-rata jumlah Goldfish crackers yang dimakan oleh anak-anak yang menonton kartun dengan iklan lebih tinggi?

## 10.10 (cont'd)

*Computer Anxiety Rating Scale (CARS)* mengukur tingkat kecemasan terhadap komputer (*computer anxiety*), dengan skala dari 20 (tidak ada kecemasan) hingga 100 (sangat cemas). Peneliti dari Miami University menyebarkan CARS pada 172 mahasiswa bisnis. Salah satu tujuan dari penelitian tersebut adalah menentukan apakah terdapat perbedaan tingkat kecemasan komputer yang dirasakan oleh mahasiswa bisnis pria dan wanita. Mereka menemukan data berikut:

	<b>Pria</b>	<b>Wanita</b>
$\bar{X}$	40,26	36,85
s	13,35	9,42
n	100	72

## 10.10

Dengan tingkat signifikansi 0.05, apakah terdapat bukti bahwa kecemasan komputer yang dirasakan oleh mahasiswa bisnis wanita berbeda dari yang dirasakan oleh mahasiswa bisnis pria?

# 10.13

Sebuah cabang Bank Mandiri yang terletak di area komersial ingin meningkatkan pelayanannya dengan memperpendek waktu tunggu. Berikut adalah data waktu tunggu untuk dilayani pada cabang Bank Mandiri tersebut (data diambil dari sampel 15 orang nasabah):

<b>Waktu tunggu(Bank 1)</b>	
Rata-rata	4,29
Standar Deviasi	1,64

# 10.13

Misalkan terdapat cabang lain yang terletak di area perumahan juga ingin memperpendek waktu tunggu pelayanannya. Pada bank ini dipilih juga sebuah sampel yang terdiri dari 15 orang nasabah, dengan hasil sebagai berikut:

<b>Waktu Tunggu (Bank 2)</b>	
Rata-rata	7,11
Standar Deviasi	2,08

## 10.13

Dengan mengasumsikan bahwa variansi populasi kedua bank tersebut tidak sama, apakah terdapat bukti yang menunjukkan rata-rata waktu menunggu di kedua cabang tersebut berbeda? (Gunakan  $\alpha = 0,05$ )



## 10.16 (cont'd)

Apakah anak-anak menggunakan telepon selular? Sepertinya demikian, menurut penelitian baru-baru ini, pengguna telepon selular berusia dibawah 12 tahun rata-rata melakukan 137 panggilan telepon per bulan. Cukup tinggi, jika dibandingkan dengan 231 panggilan telepon per bulan yang dilakukan oleh pengguna telepon selular berusia 13 hingga 17 tahun.

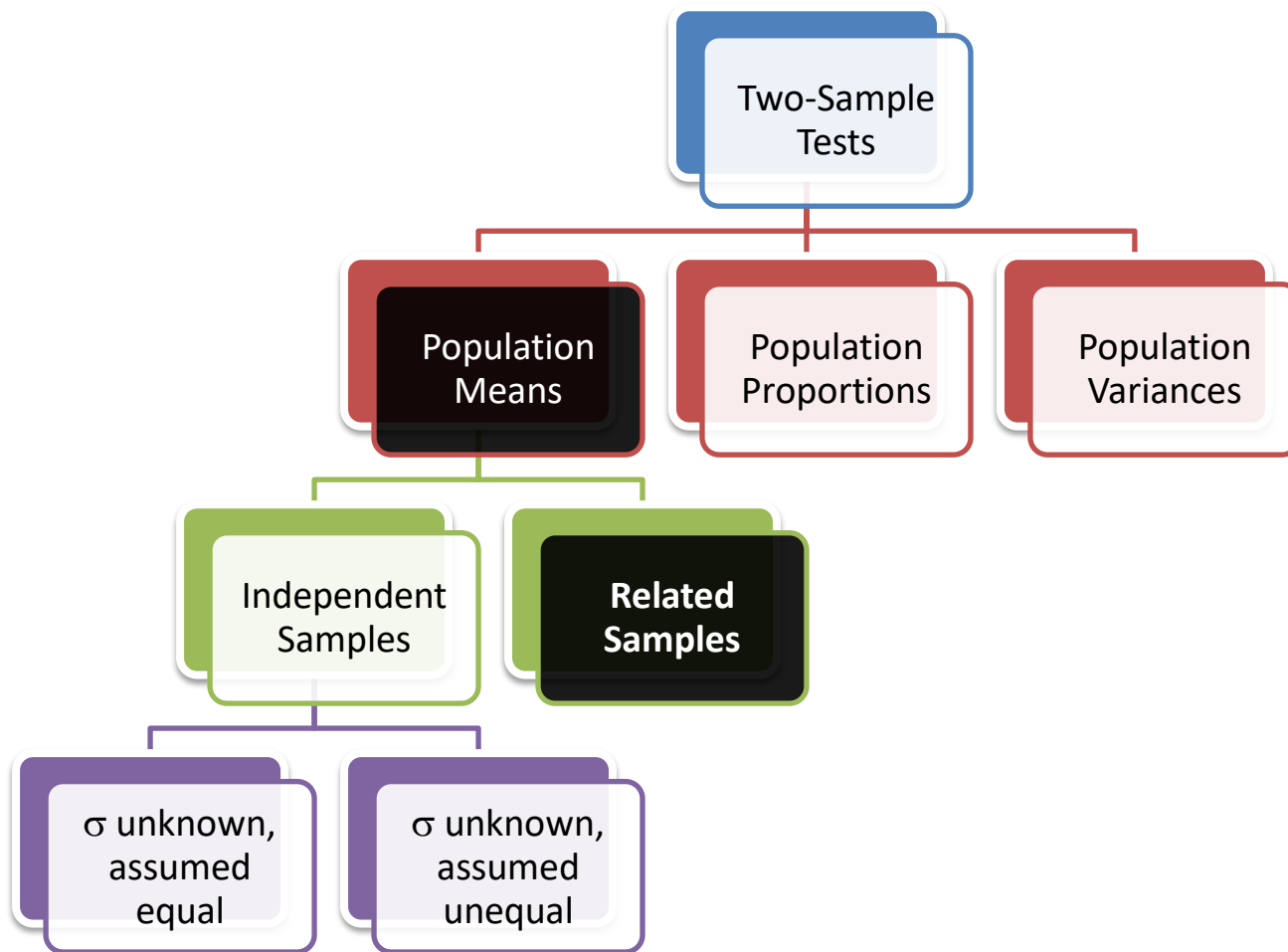
## 10.16 (cont'd)

Misalkan hasil tersebut diambil dari sampel 50 orang pengguna telepon selular untuk setiap grup pengguna dan simpangan baku sampel pengguna telepon selular berusia dibawah 12 tahun adalah 51,7 panggilan telepon per bulan dan simpangan baku sampel pengguna telepon selular berusia 13 hingga 17 tahun adalah 67,6 panggilan telepon per bulan.

## 10.16

Dengan mengasumsikan bahwa variansi populasi dari pengguna telepon selular adalah sama, adakah bukti yang menunjukkan bahwa terdapat perbedaan rata-rata penggunaan telepon selular antara kelompok usia dibawah 12 tahun dan kelompok usia 13 hingga 17 tahun? (Gunakan tingkat signifikansi 0,05.)

# Two-Sample Tests



# Related Populations

## The Paired Difference Test

Related samples

Tests Means of 2 **Related** Populations

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

$$D_i = X_{1i} - X_{2i}$$

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

# Related Populations

## The Paired Difference Test

*(continued)*

Related samples

The  $i^{\text{th}}$  paired difference is  $D_i$ , where

$$D_i = X_{1i} - X_{2i}$$

The point estimate for the paired Difference population mean  $\mu_D$  is  $\bar{D}$ :

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$$

The sample standard deviation is  $S_D$

$$S_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}}$$

$n$  is the number of pairs in the paired sample

# The Paired Difference Test: Finding $t_{\text{STAT}}$

Related samples

- The test statistic for  $\mu_D$  is:

$$t_{\text{STAT}} = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$$

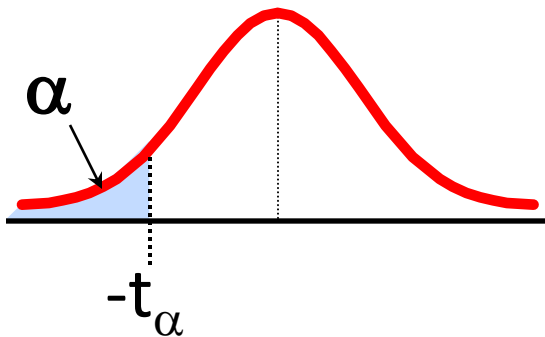
- Where  $t_{\text{STAT}}$  has  **$n - 1$  d.f.**

# The Paired Difference Test: Possible Hypotheses

## Paired Samples

Lower-tail test:

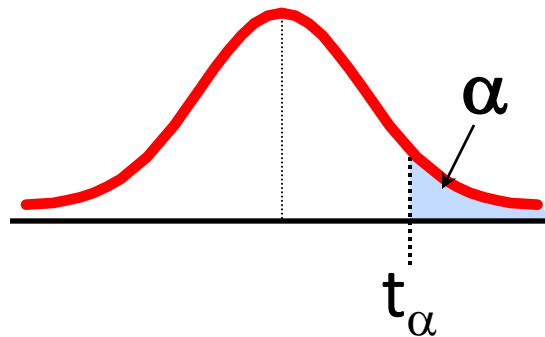
$$H_0: \mu_D \geq 0$$
$$H_1: \mu_D < 0$$



Reject  $H_0$  if  $t_{\text{STAT}} < -t_\alpha$

Upper-tail test:

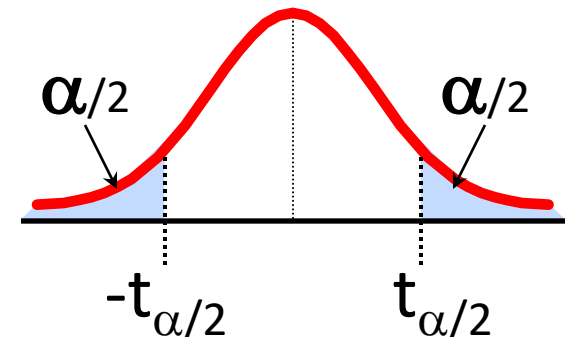
$$H_0: \mu_D \leq 0$$
$$H_1: \mu_D > 0$$



Reject  $H_0$  if  $t_{\text{STAT}} > t_\alpha$

Two-tail test:

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



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

Where  $t_{\text{STAT}}$  has  $n - 1$  d.f.



# The Paired Difference Confidence Interval

The confidence interval for  $\mu_D$  is

Related samples

$$\bar{D} \pm t_{\alpha/2} \frac{S_D}{\sqrt{n}}$$

where  $S_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}}$

# Paired Difference Test: Example

Misalkan anda mengirimkan karyawan anda untuk mengikuti pelatihan “customer service”. Apakah pelatihan ini mempengaruhi jumlah keluhan yang masuk? Anda mengumpulkan data sebagai berikut:

<u>Karyawan</u>	<u>Jumlah Keluhan:</u>	
	<u>Sebelum</u>	<u>Sesudah</u>
C.B.	6	4
T.F.	20	6
M.H.	3	2
R.K.	0	0
M.O.	4	0

# Paired Difference Test: Example

Misalkan anda mengirimkan karyawan anda untuk mengikuti pelatihan “customer service”. Apakah pelatihan ini mempengaruhi jumlah keluhan yang masuk? Anda mengumpulkan data sebagai berikut:

<u>Karyawan</u>	<u>Jumlah Keluhan:</u>		<u>(2) - (1)</u> <u>Selisih, <math>D_i</math></u>
	<u>Sebelum (1)</u>	<u>Sesudah (2)</u>	
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
			<u>-21</u>

$$\bar{D} = \frac{\sum D_i}{n}$$

$$= -4.2$$

$$S_D = \sqrt{\frac{\sum (D_i - \bar{D})^2}{n - 1}}$$

$$= 5.67$$

# Paired Difference Test: Solution

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

$$\begin{aligned} H_0: \mu_D &= 0 \\ H_1: \mu_D &\neq 0 \end{aligned}$$

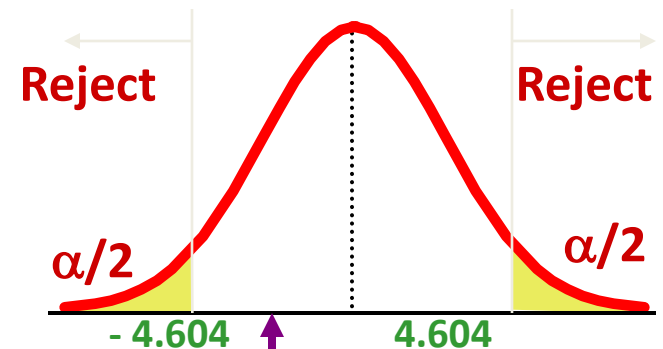
$$\alpha = .01 \quad \bar{D} = -4.2$$

$$t_{0.005} = \pm 4.604$$

$$\text{d.f.} = n - 1 = 4$$

Test Statistic:

$$t_{\text{STAT}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}} = \frac{-4.2 - 0}{5.67 / \sqrt{5}} = -1.66$$



Decision: **Do not reject  $H_0$**

( $t_{\text{stat}}$  is not in the reject region)

Conclusion: **There is insufficient evidence there is significant change in the number of complaints.**

**EXERCISE**

## 10.20 (cont'd)

Sembilan ahli memberi penilaian pada dua merek kopi Kolombia dalam sebuah percobaan *taste-testing*. Sebuah rating berskala 7 (1 = sangat tidak memuaskan, 7 = sangat memuaskan) diberikan untuk empat karakteristik: rasa, aroma, *richness*, dan keasaman. Data berikut menunjukkan total rating dari keempat karakteristik tersebut.

Merek		
Ahli	A	B
C.C.	24	26
S.E.	27	27
E.G.	19	22
B.L.	24	27
C.M.	22	25
C.N.	26	27
G.N.	27	26
R.M.	25	27
P.V.	22	23

## 10.20

Dengan tingkat signifikansi 0,05, adakah bukti yang menunjukkan bahwa rata-rata rating antara kedua merek tersebut berbeda?

## 10.24 (cont'd)

Kanker plasma darah (Multiple myeloma), dapat diketahui dari peningkatan pembentukan pembuluh darah (angiogenesis) pada sumsum tulang yang merupakan faktor penentu keberlangsungan hidup penderitanya. Salah satu pengobatan untuk penyakit ini adalah transplantasi sel induk (stem cell). Berikut merupakan data kepadatan pembuluh darah pada sumsum tulang pada pasien yang telah menyelesaikan prosedur transplantasi sel induk (diukur dengan tes darah dan urin).



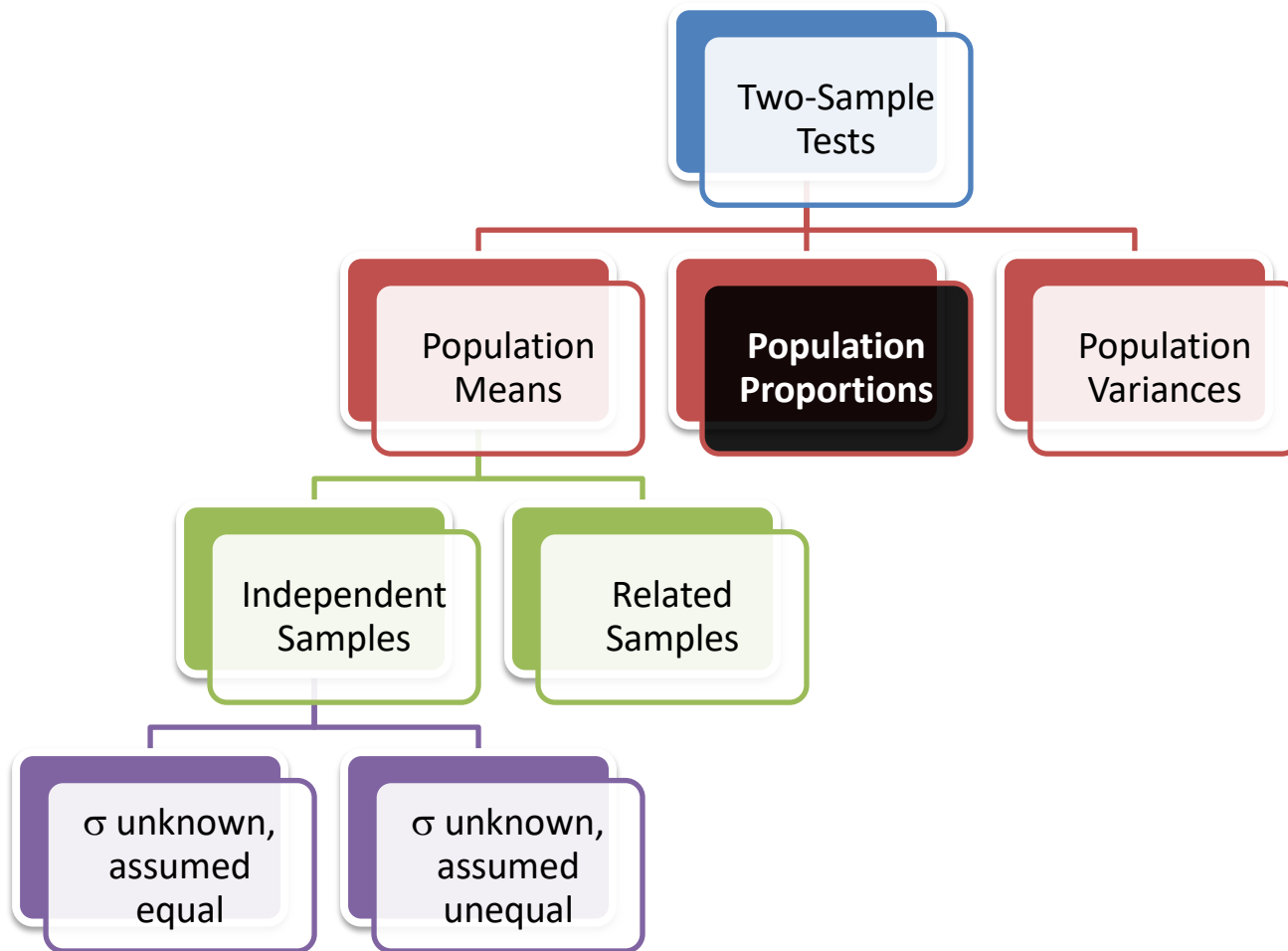
## 10.24 (cont'd)

Pasien	Sebelum	Sesudah
1	158	284
2	189	214
3	202	101
4	353	227
5	416	290
6	426	176
7	441	290

## 10.24

Pengukuran dilakukan tepat sebelum transplantasi sel induk dan ketika seluruh prosedur transplantasi sel induk selesai dilaksanakan. Pada tingkat signifikansi 0,05, apakah terdapat bukti bahwa rata-rata kepadatan pembuluh darah pada sumsum tulang sebelum transplantasi sel induk lebih tinggi jika dibandingkan dengan setelah transplantasi?

# Two-Sample Tests



# Two Population Proportions

Population proportions

**Goal:** test a hypothesis or form a confidence interval for the difference between two population proportions,

$$\pi_1 - \pi_2$$

The point estimate for the difference is

$$p_1 - p_2$$

# Two Population Proportions

Population  
proportions

In the null hypothesis we assume the null hypothesis is true, so we assume  $\pi_1 = \pi_2$  and pool the two sample estimates

The pooled estimate for the overall proportion is:

$$\bar{p} = \frac{\mathbf{X}_1 + \mathbf{X}_2}{\mathbf{n}_1 + \mathbf{n}_2}$$

where  $X_1$  and  $X_2$  are the number of items of interest in samples 1 and 2

# Two Population Proportions

(continued)

Population  
proportions

The test statistic for  
 $\pi_1 - \pi_2$  is a Z statistic:

$$Z_{STAT} = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}(1 - \bar{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where  $\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$  ,  $p_1 = \frac{X_1}{n_1}$  ,  $p_2 = \frac{X_2}{n_2}$

# Hypothesis Tests for Two Population Proportions

Population proportions

Lower-tail test:

$$H_0: \pi_1 \geq \pi_2$$

$$H_1: \pi_1 < \pi_2$$

i.e.,

$$H_0: \pi_1 - \pi_2 \geq 0$$

$$H_1: \pi_1 - \pi_2 < 0$$

Upper-tail test:

$$H_0: \pi_1 \leq \pi_2$$

$$H_1: \pi_1 > \pi_2$$

i.e.,

$$H_0: \pi_1 - \pi_2 \leq 0$$

$$H_1: \pi_1 - \pi_2 > 0$$

Two-tail test:

$$H_0: \pi_1 = \pi_2$$

$$H_1: \pi_1 \neq \pi_2$$

i.e.,

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 \neq 0$$

# Hypothesis Tests for Two Population Proportions

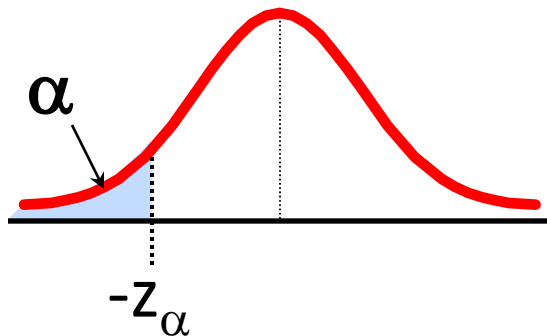
(continued)

Population proportions

Lower-tail test:

$$H_0: \pi_1 - \pi_2 \geq 0$$

$$H_1: \pi_1 - \pi_2 < 0$$

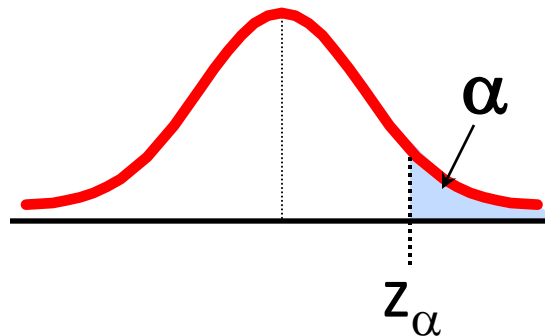


Reject  $H_0$  if  $Z_{\text{STAT}} < -Z_\alpha$

Upper-tail test:

$$H_0: \pi_1 - \pi_2 \leq 0$$

$$H_1: \pi_1 - \pi_2 > 0$$

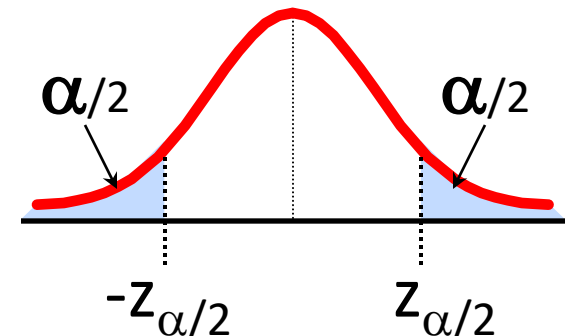


Reject  $H_0$  if  $Z_{\text{STAT}} > Z_\alpha$

Two-tail test:

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 \neq 0$$



Reject  $H_0$  if  $Z_{\text{STAT}} < -Z_{\alpha/2}$   
or  $Z_{\text{STAT}} > Z_{\alpha/2}$



# Hypothesis Test 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 35 of 50 women indicated they would vote Yes
- Test at the .05 level of significance



# Hypothesis Test Example: Two population Proportions

*(continued)*

- The hypothesis test is:

$H_0: \pi_1 - \pi_2 = 0$  (the two proportions are equal)

$H_1: \pi_1 - \pi_2 \neq 0$  (there is a significant difference between proportions)

- The sample proportions are:

- Men:  $p_1 = 36/72 = 0.50$

- Women:  $p_2 = 35/50 = 0.70$

- The pooled estimate for the overall proportion is:

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{36 + 35}{72 + 50} = \frac{71}{122} = 0.582$$

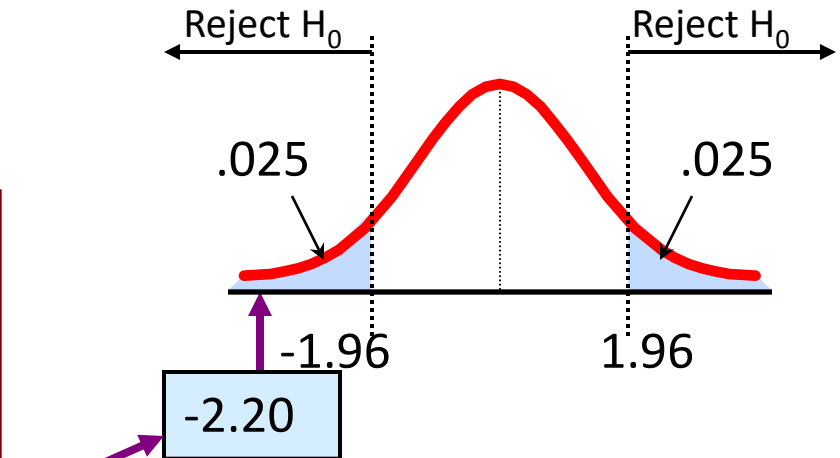
# Hypothesis Test Example: Two population Proportions

(continued)

The test statistic for  $\pi_1 - \pi_2$  is:

$$\begin{aligned} z_{STAT} &= \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\ &= \frac{(.50 - .70) - (0)}{\sqrt{.582(1 - .582)\left(\frac{1}{72} + \frac{1}{50}\right)}} = -2.20 \end{aligned}$$

**Critical Values =  $\pm 1.96$**   
**For  $\alpha = .05$**



**Decision: Reject  $H_0$**

**Conclusion: There is evidence of a difference in proportions who will vote yes between men and women.**

# Confidence Interval for Two Population Proportions

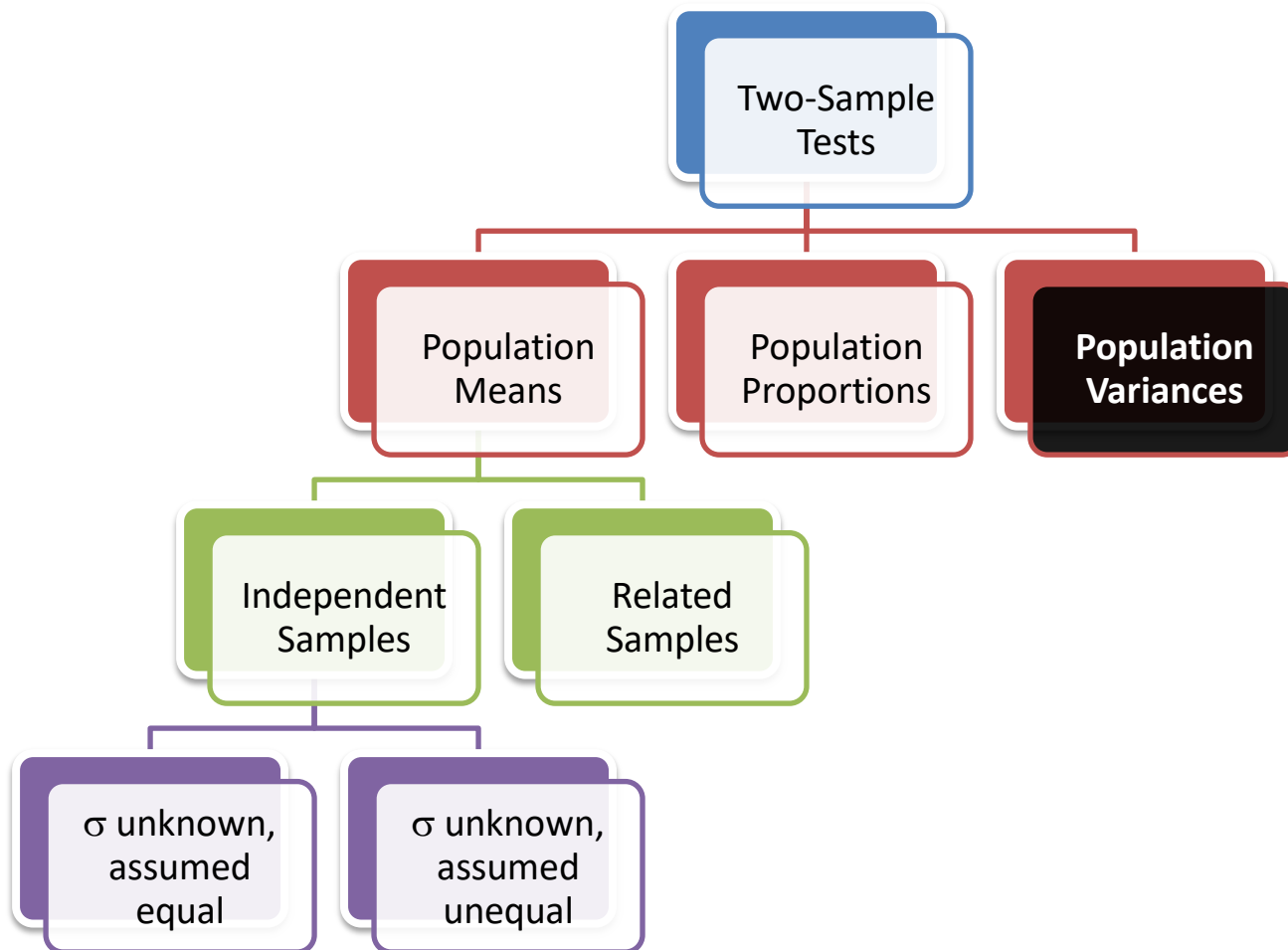
Population proportions

The confidence interval for

$\pi_1 - \pi_2$  is:

$$(\mathbf{p}_1 - \mathbf{p}_2) \pm \mathbf{Z}_{\alpha/2} \sqrt{\frac{\mathbf{p}_1(1 - \mathbf{p}_1)}{\mathbf{n}_1} + \frac{\mathbf{p}_2(1 - \mathbf{p}_2)}{\mathbf{n}_2}}$$

# Two-Sample Tests



# Testing for the Ratio Of Two Population Variances

Tests for Two Population Variances

F test statistic

Hypotheses

$F_{STAT}$

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$H_0: \sigma_1^2 \leq \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$\frac{S_1^2}{S_2^2}$$

Where:

$S_1^2$  = Variance of sample 1 (the larger sample variance)

$n_1$  = sample size of sample 1

$S_2^2$  = Variance of sample 2 (the smaller sample variance)

$n_2$  = sample size of sample 2

$n_1 - 1$  = numerator degrees of freedom

$n_2 - 1$  = denominator degrees of freedom

# The F Distribution

- The F critical value is found from the F table
- There are two degrees of freedom required: numerator and denominator
- The larger sample variance is always the numerator

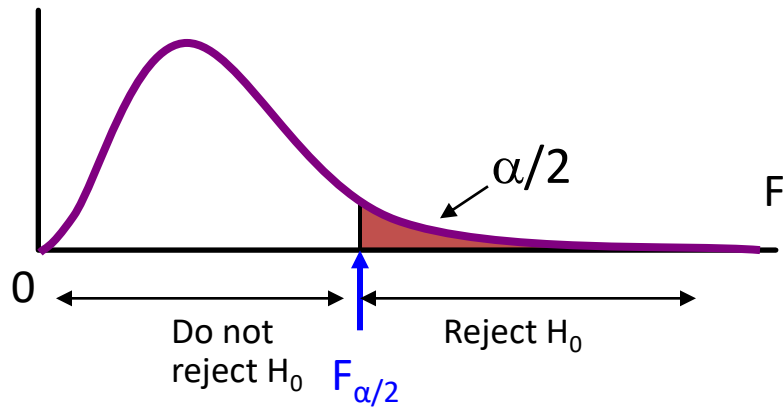
- When  $F_{STAT} = \frac{S_1^2}{S_2^2}$   $df_1 = n_1 - 1$  ;  $df_2 = n_2 - 1$

- In the F table,
  - numerator degrees of freedom determine the column
  - denominator degrees of freedom determine the row

# Finding the Rejection Region

$$H_0: \sigma_1^2 = \sigma_2^2$$

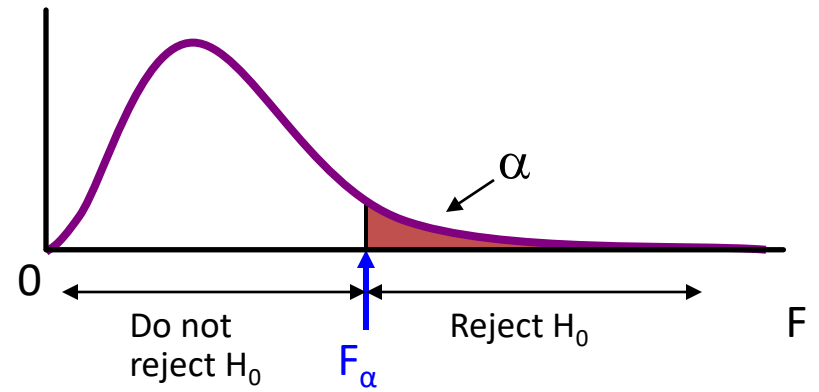
$$H_1: \sigma_1^2 \neq \sigma_2^2$$



Reject  $H_0$  if  $F_{STAT} > F_{\alpha/2}$

$$H_0: \sigma_1^2 \leq \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$



Reject  $H_0$  if  $F_{STAT} > F_{\alpha}$



# F Test: An Example

Anda adalah seorang analis keuangan dan ingin membandingkan hasil dividen antara saham yang terdaftar di NYSE & NASDAQ? Anda mengumpulkan data sebagai berikut:

	<u>NYSE</u>	<u>NASDAQ</u>
<b>Jumlah</b>	<b>21</b>	<b>25</b>
<b>Rata-rata sampel</b>	<b>3.27</b>	<b>2.53</b>
<b>Std dev sampel</b>	<b>1.30</b>	<b>1.16</b>



Apakah terdapat perbedaan variansi antara hasil dividen pada NYSE & NASDAQ pada tingkat signifikansi 0,05?

# F Test: Example Solution

- Form the hypothesis test:

$$H_0: \sigma_1^2 = \sigma_2^2$$

(there is no difference between variances)

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

(there is a difference between variances)

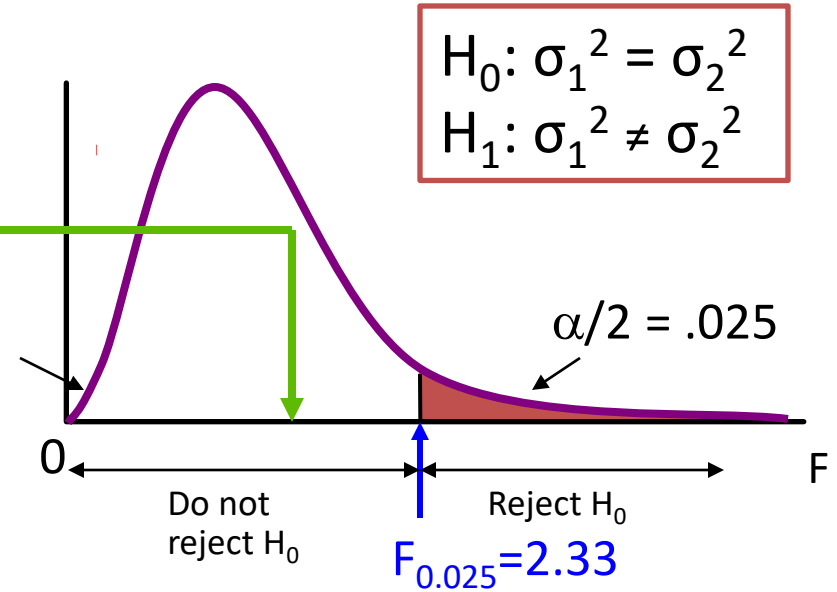
- Find the F critical value for  $\alpha = 0.05$ :
- Numerator d.f. =  $n_1 - 1 = 21 - 1 = 20$
- Denominator d.f. =  $n_2 - 1 = 25 - 1 = 24$
- $F_{\alpha/2} = F_{.025, 20, 24} = 2.33$

# F Test: Example Solution

(continued)

- The test statistic is:

$$F_{STAT} = \frac{S_1^2}{S_2^2} = \frac{1.30^2}{1.16^2} = 1.256$$



- $F_{STAT} = 1.256$  is not in the rejection region, so we **do not reject  $H_0$**
- Conclusion:** There is insufficient evidence of a difference in variances at  $\alpha = .05$

**EXERCISE**

## 10.30 (cont'd)

Apakah tahun ini diperlukan usaha lebih untuk keluar dari sebuah *mailing list* dari tahun sebelumnya? Sebuah penelitian dari 100 peritel online besar menunjukkan data berikut:

PERLU TIGA ATAU LEBIH KLIK SEBELUM KELUAR		
TAHUN	Ya	Tidak
2009	39	61
2008	7	93

## 10.30

- a. Tentukan hipotesis kosong dan hipotesis alternatif untuk mengetahui apakah diperlukan usaha lebih untuk keluar dari sebuah *mailing list* jika dibandingkan dengan tahun sebelumnya.
- b. Lakukan uji hipotesis untuk poin (a), dengan menggunakan tingkat signifikansi 0,05.
- c. Apakah hasil dari poin (b) sesuai dengan klaim bahwa diperlukan usaha lebih untuk keluar dari sebuah *mailing list* jika dibandingkan dengan tahun sebelumnya?

## 10.34 (cont'd)

Bagaimana perasaan orang Amerika mengenai iklan di halaman web? Sebuah survey yang dilakukan pada 1.000 pengguna internet dewasa diketahui bahwa 670 orang menentang adanya iklan di halaman web. Misalkan sebuah survei lain pada 1.000 orang pengguna internet berusia 12–17 tahun menemukan bahwa 510 orang menentang adanya iklan di halaman web. Dengan menggunakan tingkat signifikansi 0,05, adakah bukti bahwa terdapat perbedaan proporsi antara pengguna internet dewasa dan pengguna internet berusia 12–17 tahun yang menentang iklan?

## 10.46 (cont'd)

*Computer Anxiety Rating Scale (CARS)* mengukur tingkat kecemasan terhadap komputer (*computer anxiety*), dengan skala dari 20 (tidak ada kecemasan) hingga 100 (sangat cemas). Peneliti dari Miami University menyebarjab CARS pada 172 mahasiswa bisnis. Salah satu tujuan dari penelitian tersebut adalah menentukan apakah terdapat perbedaan tingkat kecemasan komputer yang dirasakan oleh mahasiswa bisnis pria dan wanita. Mereka menemukan data berikut:

	<b>Pria</b>	<b>Wanita</b>
X	40.26	36.85
S	13.35	9.42
n	100	72



# 10.46

- a. Dengan tingkat signifikansi 0,05, adakah bukti yang menunjukkan perbedaan sebaran (variability) kecemasan komputer (computer anxiety) yang dialami pria dan wanita?
- b. Asumsi apa yang anda perlukan tentang kedua populasi tersebut untuk dapat menggunakan uji  $F$ ?
- c. Berdasarkan poin (a) dan (b), uji  $t$  manakah yang seharusnya anda gunakan untuk menguji apakah ada perbedaan yang signifikan antara kecemasan komputer yang dirasakan oleh wanita dan pria?

**THANK YOU**