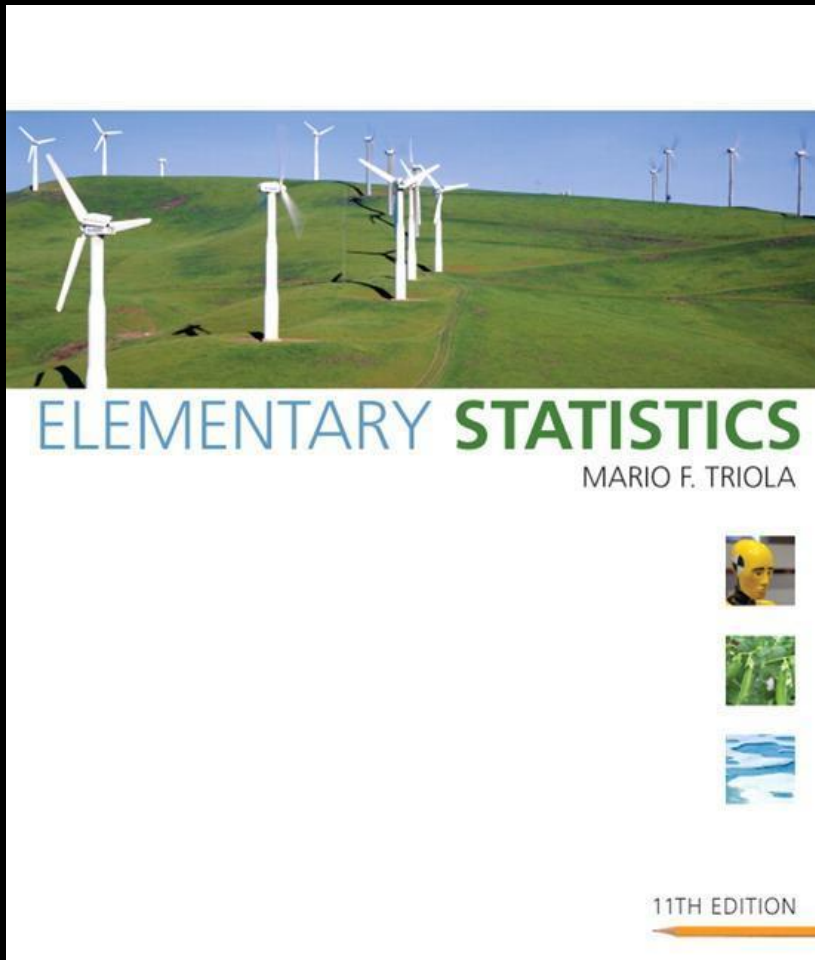# Lecture Slides

Elementary Statistics
Eleventh Edition

and the Triola Statistics Series

by Mario F. Triola

ELEMENTARY **STATISTICS**

MARIO F. TRIOLA

11TH EDITION

PEARSO

Chapter 12
Analysis of Variance

Section 12-1
Review and Preview

Review

In chapter 9, we introduced methods for comparing the means from two independent samples.

Preview

❖ **Analysis of variance (ANOVA) is a method for testing the hypothesis that three or more population means are equal.**

❖ **For example:**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$$

$$H_1: \text{At least one mean is different}$$

ANOVA Methods
Require the *F*-Distribution

1. The *F*-distribution is not symmetric; it is skewed to the right.

2. The values of *F* can be 0 or positive; they cannot be negative.

3. There is a different *F*-distribution for each pair of degrees of freedom for the numerator and denominator.

Critical values of *F* are given in **Table A-5**

F-Distribution

Not symmetric
(skewed to the right)

α

0   1.0   *F*

Nonnegative
values only
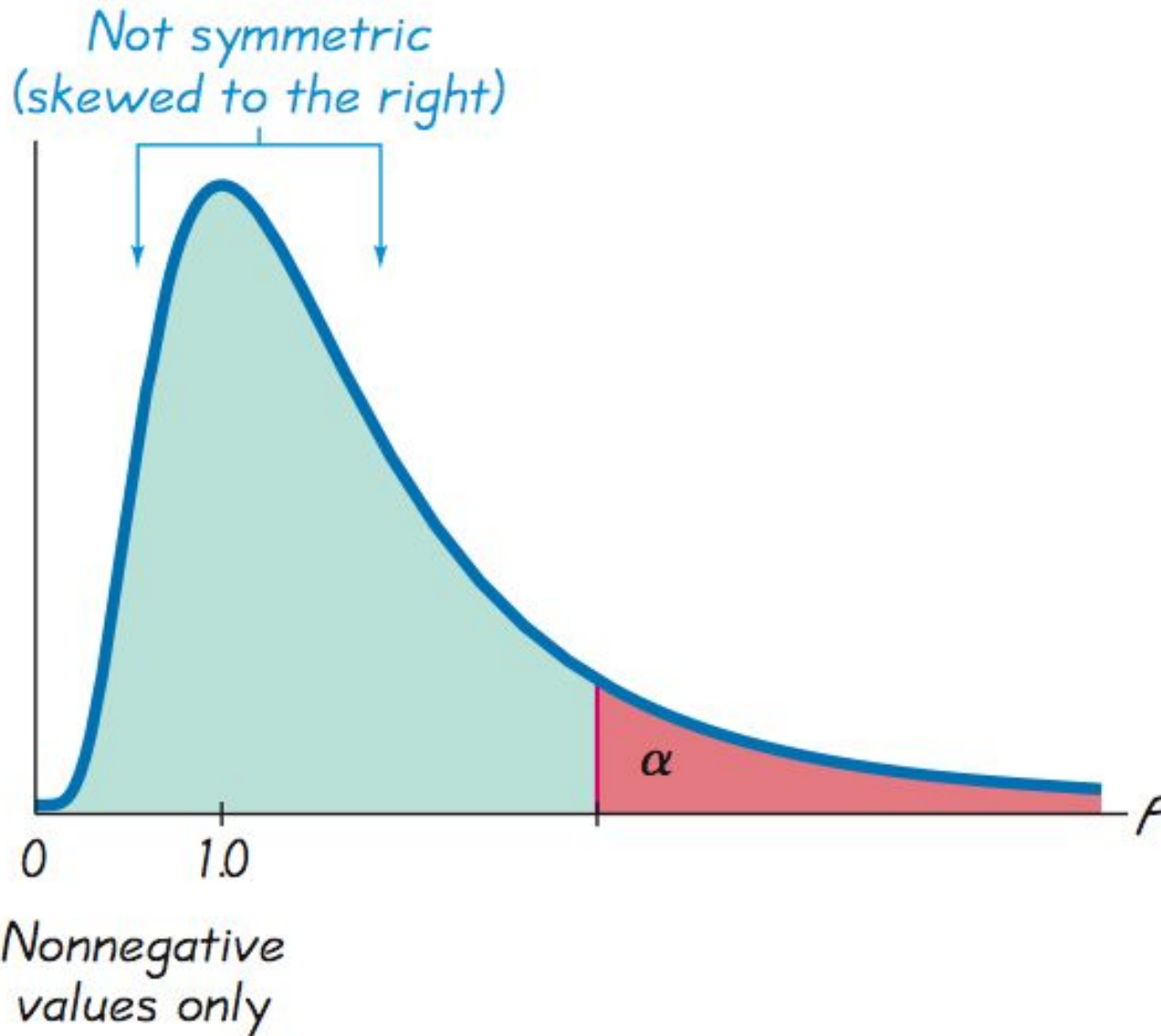
Section 12-2
One-Way ANOVA

Key Concept

This section introduces the method of **one-way analysis of variance**, which is used for tests of hypotheses that three or more population means are all equal.
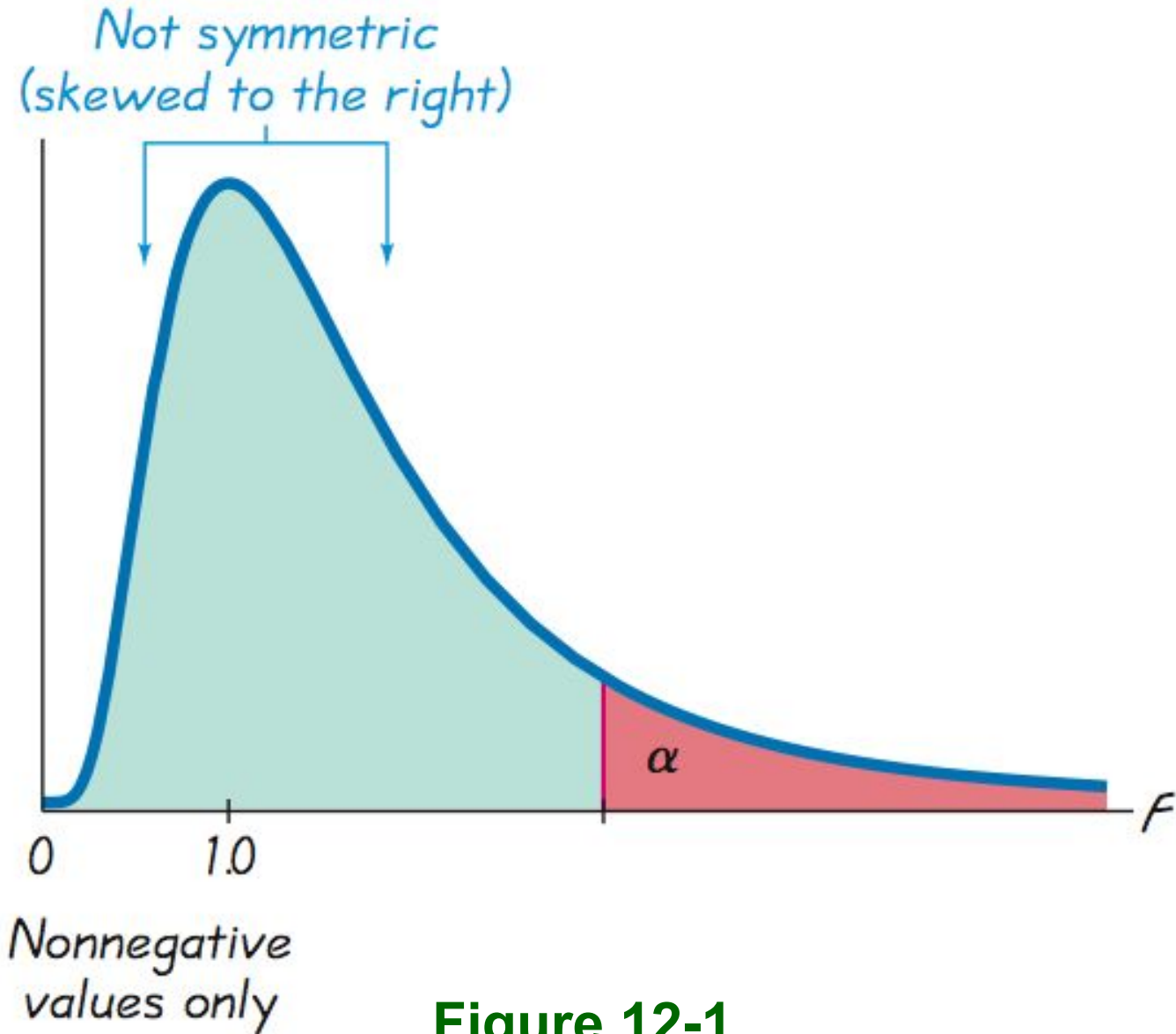
F - distribution

Figure 12-1

12.1 - *

One-Way ANOVA

An Approach to Understanding ANOVA

1. Understand that a small P-value (such as 0.05 or less) leads to rejection of the null hypothesis of equal means.

    With a large P-value (such as greater than 0.05), fail to reject the null hypothesis of equal means.

2. Develop an understanding of the underlying rationale by studying the examples in this section.

Part 1: Basics of One-Way Analysis of Variance

Definition

One-way analysis of variance (ANOVA) is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with *one* treatment (or factor), which is a characteristic that allows us to distinguish the different populations from one another.

One-Way ANOVA

Requirements

1. The populations have approximately normal distributions.

2. The populations have the same variance σ^2 (or standard deviation σ).

3. The samples are simple random samples.

4. The samples are independent of each other.

5. The different samples are from populations that are categorized in only one way.

Procedure for testing
$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots$

1. Use STATDISK, Minitab, Excel, or a TI-83/84 Calculator to obtain results.

2. Identify the P-value from the display.

3. Form a conclusion based on these criteria:

Procedure for testing H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots$

If the P-value $\leq \alpha$, reject the null hypothesis of equal means and conclude that at least one of the population means is different from the others.

If the P-value $> \alpha$, fail to reject the null hypothesis of equal means.

Caution

When we conclude that there is sufficient evidence to reject the claim of equal population means, we cannot conclude from ANOVA that any particular mean is different from the others.

There are several other tests that can be used to identify the specific means that are different, and some of them are discussed in Part 2 of this section.

Example:

Use the chest deceleration measurements listed in Table 12-1 and a significance level of α = 0.05 to test the claim that the three samples come from populations with means that are all equal.

Table 12-1	Chest Deceleration Measurements (in g) from Car Crash Tests										
Small Cars	44	43	44	54	38	43	42	45	44	50	→ $\bar{x}$ = 44.7 g
Medium Cars	41	49	43	41	47	42	37	43	44	34	→ $\bar{x}$ = 42.1 g
Large Cars	32	37	38	45	37	33	38	45	43	42	→ $\bar{x}$ = 39.0 g

Example:

Requirements are satisfied: distributions are approximately normal (normal quantile plots); population variances appear to be about the same; simple random samples; independent samples, not matched; categorized according to a single factor of size

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1:   **At least one of the means is different from the others**

 significance level is $\alpha = 0.05$

Example:

Step 1: Use technology to obtain ANOVA results

STATDISK

Source:	DF:	SS:	MS:	Test Stat, F:	Critical F:	P-Value:
Treatment:	2	162.866667	81.433333	4.094413	3.354128	0.027986
Error:	27	537.00	19.888889			
Total:	29	699.866667				

Reject the Null Hypothesis
Reject equality of means

MINITAB

One-way ANOVA: Small, Medium, Large

Source	DF	SS	MS	F	P
Factor	2	162.9	81.4	4.09	0.028
Error	27	537.0	19.9		
Total	29	699.9			

EXCEL

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	162.8667	2	81.43333	4.094413	0.027986	3.354131
Within Groups	537	27	19.88889			
Total	699.8667	29				

Example:

Step 1: Use technology to obtain ANOVA results

EXCEL

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	162.8667	2	81.43333	4.094413	0.027986	3.354131
Within Groups	537	27	19.88889			
Total	699.8667	29				

TI-83/84 PLUS

```
One-way ANOVA
 F=4.094413408
 P=.0279863448
 Factor
  df=2
  SS=162.866667
↓ MS=81.4333333
```

```
One-way ANOVA
↑ MS=81.4333333
 Error
  df=27
  SS=537
  MS=19.8888889
 Sxp=4.45969605
```

Example:

Step 2: Displays all show *P*-value = 0.28 when rounded

Step 3: Because the P-value of 0.028 is less than the significance level of α = 0.05, we reject the null hypothesis of equal means.

There is sufficient evidence to warrant rejection of the claim that the three samples come from populations with means that are all equal.

Example:

Based on the samples of measurements listed in Table 12-1, we conclude that those values come from populations having means that are not all the same. On the basis of this ANOVA test, we cannot conclude that any particular mean is different from the others, but we can informally note that the sample mean is smallest for the large cars. Because small measurements correspond to less trauma experienced by the crash test dummies, it appears that the large cars are safest, but this conclusion is not formally justified by this ANOVA test.
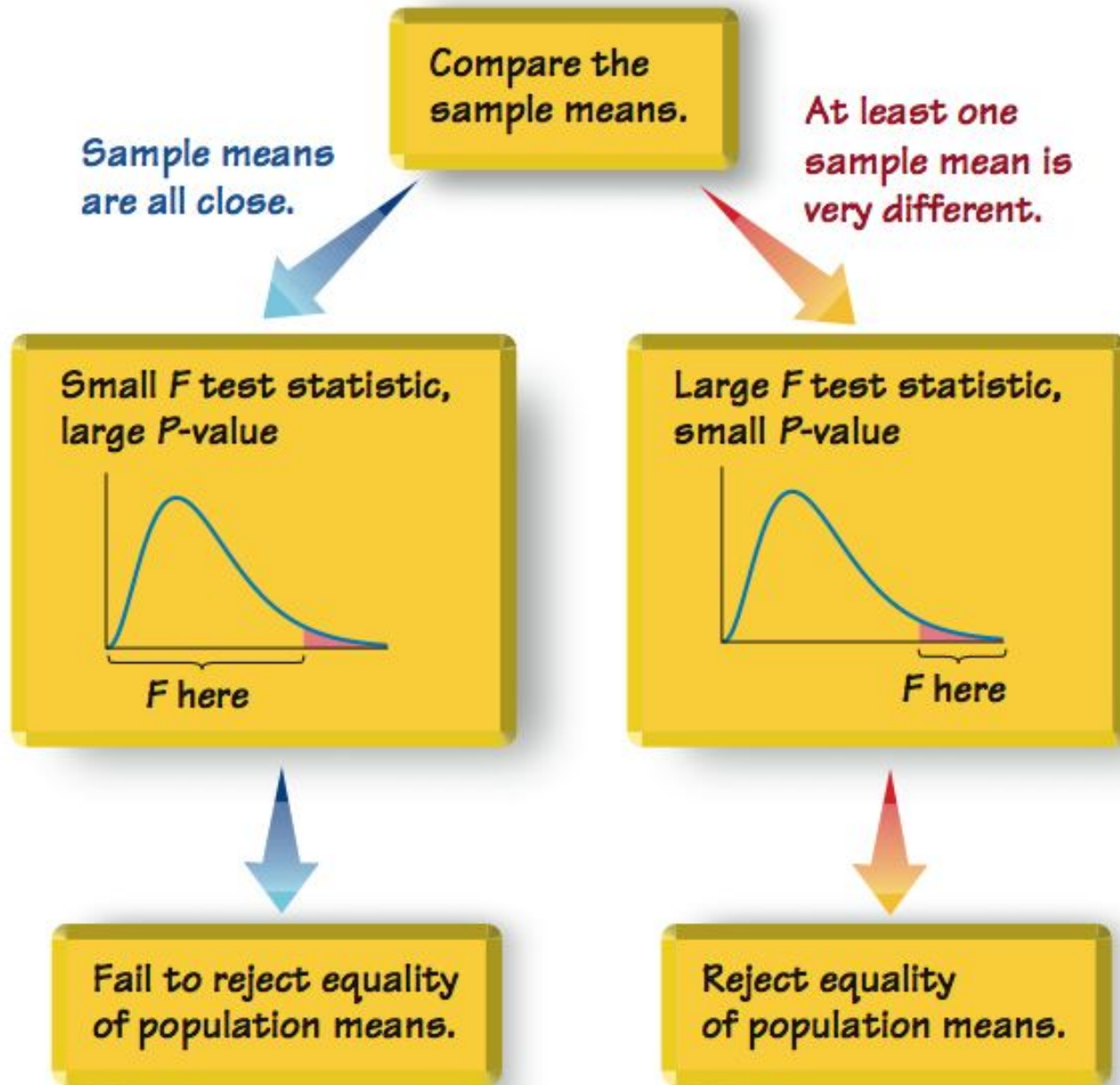
P-Value and Test Statistic

Larger values of the test statistic result in *smaller P*-values, so the ANOVA test is right-tailed. Figure 12-2 shows the relationship between the F test statistic and the *P*-value. Assuming that the populations have the same variance σ^2 (as required for the test), the F test statistic is the ratio of these two estimates of σ^2: (1) variation *between* samples (based on variation among sample means); and (2) variation *within* samples (based on the sample variances).

Test Statistic for One-Way ANOVA

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

Relationship Between *F* Test Statistic / *P*-Value

Compare the sample means.

Sample means are all close.

At least one sample mean is very different.

Small *F* test statistic, large *P*-value

F here

Large *F* test statistic, small *P*-value

F here

Fail to reject equality of population means.

Reject equality of population means.

Caution

When testing for equality of three or more populations, use analysis of variance. Do not use multiple hypothesis tests with two samples at a time.

Part 2: Calculations and Identifying Means That Are Different

ANOVA
Fundamental Concepts

Estimate the common value of σ^2:

1. The **variance <u>between</u> samples** (also called **variation due to treatment**) is an estimate of the common population variance σ^2 that is based on the variability among the sample <u>means</u>.

2. The **variance <u>within</u> samples** (also called **variation due to error**) is an estimate of the common population variance σ^2 based on the sample <u>variances</u>.

ANOVA
Fundamental Concepts

Test Statistic for One-Way ANOVA

$$F = \frac{\text{variance } \textbf{between} \text{ samples}}{\text{variance } \textbf{within} \text{ samples}}$$

An excessively large F test statistic is evidence against equal population means.

Calculations with Equal Sample Sizes

❖ **Variance between samples = $n\ s_{\bar{x}}^2$**

where $s_{\bar{x}}^2$ = variance of sample means

❖ **Variance within samples = s_p^2**

where s_p^2 = pooled variance (or the mean of the sample variances)

Example: Sample Calculations

Table 12-2   Effect of a Mean on the *F* Test Statistic

	A	**add 10**			**B**	
Sample 1	Sample 2	Sample 3		Sample 1	Sample 2	Sample 3

A Sample 1	A Sample 2	A Sample 3	B Sample 1	B Sample 2	B Sample 3
7	6	4	17	6	4
3	5	7	13	5	7
6	5	6	16	5	6
6	8	7	16	8	7

A
$n_1 = 4$ $\bar{x}_1 = 5.5$ $s_1^2 = 3.0$
$n_2 = 4$ $\bar{x}_2 = 6.0$ $s_2^2 = 2.0$
$n_3 = 4$ $\bar{x}_3 = 6.0$ $s_3^2 = 2.0$

B
$n_1 = 4$ $\bar{x}_1 = 15.5$ $s_1^2 = 3.0$
$n_2 = 4$ $\bar{x}_2 = 6.0$ $s_2^2 = 2.0$
$n_3 = 4$ $\bar{x}_3 = 6.0$ $s_3^2 = 2.0$

Variance between samples

A: $ns_{\bar{x}}^2 = 4\,(0.0833) = 0.3332$

B: $ns_{\bar{x}}^2 = 4\,(30.0833) = 120.3332$

Variance within samples

A: $s_p^2 = \dfrac{3.0 + 2.0 + 2.0}{3} = 2.3333$

B: $s_p^2 = \dfrac{3.0 + 2.0 + 2.0}{3} = 2.3333$

F test statistic

A: $F = \dfrac{ns_{\bar{x}}^2}{s_p^2} = \dfrac{0.3332}{2.3333} = \mathbf{0.1428}$

B: $F = \dfrac{ns_{\bar{x}}^2}{s_p^2} = \dfrac{120.3332}{2.3333} = \mathbf{51.5721}$

P-value (found from Excel)

A: *P*-value = 0.8688

B: *P*-value = 0.0000118

Example: Sample Calculations

Use Table 12-2 to calculate the variance between samples, variance within samples, and the *F* test statistic.

1. Find the variance ***between*** samples = $ns_{\bar{x}}^2$.

For the means 5.5, 6.0 & 6.0, the sample variance is

$$s_{\bar{x}}^2 = 0.0833$$

$$ns_{\bar{x}}^2 = 4 \times 0.0833 = 0.3332$$

2. Estimate the variance within samples by calculating the mean of the sample variances. .

$$s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$$

Example: Sample Calculations

Use Table 12-2 to calculate the variance between samples, variance within samples, and the *F* test statistic.

3. **Evaluate the *F* test statistic**

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

$$F = \frac{0.3332}{2.3333} = 0.1428$$

Critical Value of F

❖ **Right-tailed test**

❖ **Degree of freedom with k samples of the same size n**

$$\text{numerator } df = k - 1$$

$$\text{denominator } df = k(n - 1)$$

Calculations with <u>Unequal</u> Sample Sizes

$$F = \frac{\text{variance within samples}}{\text{variance between samples}} = \frac{\left[\dfrac{\Sigma n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1}\right]}{\left[\dfrac{\Sigma(n_i - 1)^2}{s_i\Sigma(n_i - 1)}\right]}$$

where $\bar{\bar{x}}$ = mean of all sample scores combined

k = number of population means being compared

n_i = number of values in the ith sample

x_i = mean of values in the ith sample

s_i = variance of values in the ith sample

Key Components of the ANOVA Method

SS(total), **or total sum of squares, is a measure of the total variation (around $\bar{\bar{x}}$) in all the sample data combined.**

Formula 12-1

$$SS\left(\text{total}\right) = \sum \left(x - \bar{\bar{x}}\right)^2$$

Key Components of the ANOVA Method

SS(treatment), also referred to as SS(factor) or SS(between groups) or SS(between samples), is a measure of the variation between the sample means.

$$SS(\text{treatment})$$

$$= n_1\left(\overline{x}_1 - \overline{\overline{x}}\right)^2 + n_2\left(\overline{x}_2 - \overline{\overline{x}}\right)^2 + \cdots + n_k\left(\overline{x}_k - \overline{\overline{x}}\right)^2$$

$$= \sum n_i\left(x_i - \overline{\overline{x}}\right)^2$$

Key Components of the ANOVA Method

SS(error), also referred to as SS(within groups) or SS(within samples), is a sum of squares representing the variability that is assumed to be common to all the populations being considered.

Formula 12-3

$$SS(error)$$

$$= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2$$

$$= \sum (n_i - 1)s_i^2$$

Key Components of the ANOVA Method

Given the previous expressions for SS(total), SS(treatment), and SS(error), the following relationship will always hold.

Formula 12-4

$$SS(total) = SS(treatment) + SS(error)$$

Mean Squares (MS)

MS(treatment) is a mean square for treatment, obtained as follows:

$$MS\text{(treatment)} = \frac{SS\text{(treatment)}}{k-1}$$

MS(error) is a mean square for error, obtained as follows:

$$MS\text{(error)} = \frac{SS\text{(error)}}{N-k}$$

N = total number of values in all samples combined

Mean Squares (MS)

MS(total) is a mean square for the total variation, obtained as follows:

Formula 12-7

$$MS(\text{total}) = \frac{SS(\text{total})}{N - 1}$$

N = total number of values in all samples combined

Test Statistic for ANOVA with <u>Unequal</u> Sample Sizes

Formula 12-8

$$F = \frac{MS\ (treatment)}{MS\ (error)}$$

❖ **Numerator** $df = k - 1$

❖ **Denominator** $df = N - k$

Identifying Means That Are Different

After conducting an analysis of variance test, we might conclude that there is sufficient evidence to reject a claim of equal population means, but we cannot conclude from ANOVA that any particular mean is different from the others.

Identifying Means That Are Different

Informal methods for comparing means

1. Use the same scale for constructing boxplots of the data sets to see if one or more of the data sets are very different from the others.

2. Construct confidence interval estimates of the means from the data sets, then compare those confidence intervals to see if one or more of them do not overlap with the others.

Bonferroni Multiple Comparison Test

Step 1. Do a separate *t* test for each pair of samples, but make the adjustments described in the following steps.

Step 2. For an estimate of the variance σ^2 that is common to all of the involved populations, use the value of MS(error).

Bonferroni Multiple Comparison Test

Step 2 (cont.) Using the value of MS(error), calculate the value of the test statistic, as shown below. (This example shows the comparison for Sample 1 and Sample 4.)

$$t = \frac{\overline{x}_1 - \overline{x}_4}{\sqrt{MS(error) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_4} \right)}}$$

Bonferroni Multiple Comparison Test

Step 2 (cont.) Change the subscripts and use another pair of samples until all of the different possible pairs of samples have been tested.

Step 3. After calculating the value of the test statistic t for a particular pair of samples, find either the critical t value or the P-value, but make the following adjustment:

Bonferroni Multiple Comparison Test

Step 3 (cont.) _P_-value: Use the test statistic _t_ with _df_ = _N_ – _k_, where _N_ is the total number of sample values and _k_ is the number of samples. Find the _P_-value the usual way, but adjust the _P_-value by multiplying it by the number of different possible pairings of two samples.

(For example, with four samples, there are six different possible pairings, so adjust the P-value by multiplying it by 6.)

Bonferroni Multiple Comparison Test

Step 3 (cont.) Critical value: When finding the critical value, adjust the significance level α by dividing it by the number of different possible pairings of two samples.

(For example, with four samples, there are six different possible pairings, so adjust the significance level by dividing it by 6.)

Example:

Using the data in Table 12-1, we concluded that there is sufficient evidence to warrant rejection of the claim of equal means. Use the Bonferroni test with a 0.05 significance level to identify which mean is different from the others.

The Bonferroni test requires a separate ttest for each different possible pair of samples. Here are the null hypotheses to be tested:

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 = \mu_3 \qquad H_0 : \mu_2 = \mu_3$$

Example:

Begin with $H_0 : \mu_1 = \mu_2$

From Table 12-1:

$$x_1 = 44.7 \quad n_1 = 10 \quad x_2 = 42.1 \quad n_2 = 10$$

MS(error) = 19.888889 (from technology)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MS(error) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}}$$

$$= \frac{44.7 - 42.1}{\sqrt{19.888889 \cdot \left(\dfrac{1}{10} + \dfrac{1}{10} \right)}} = 1.303626224$$

Example:

The number of degrees of freedom is df = $N - k$ = 30 – 3 = 27. With a test statistic of t = 1.303626244 and with df = 27, the two-tailed P-value is 0.203368, but we adjust this P-value by multiplying it by 3 (the number of different possible pairs of samples) to get a final P-value of 0.610. Because this P-value is not small (less than 0.05), we fail to reject the null hypothesis. It appears that Samples 1 and 2 do not have significantly different means.

Example:

Here are ALL the Bonferroni test results:

SPSS BONFERRONI RESULTS

Dependent Variable: Chest

Bonferroni

(I) Sample	(J) Sample	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	2.60000	1.99444	.610	-2.4907	7.6907
	3	5.70000*	1.99444	.024	.6093	10.7907
2	1	-2.60000	1.99444	.610	-7.6907	2.4907
	3	3.10000	1.99444	.395	-1.9907	8.1907
3	1	-5.70000*	1.99444	.024	-10.7907	-.6093
	2	-3.10000	1.99444	.395	-8.1907	1.9907

*. The mean difference is significant at the .05 level.

Example:

The display shows that the mean for Sample 1 (small cars) is significantly different from the mean for Sample 3 (large cars). Based on the Bonferroni test, it appears that the measurements from small cars have a mean that is significantly different from the mean for large cars.

Recap

In this section we have discussed:

❖ **One-way analysis of variance (ANOVA)**

   Calculations with <u>Equal</u> Sample Sizes

   Calculations with <u>Unequal</u> Sample Sizes

❖ **Identifying Means That Are Different**

   Bonferroni Multiple Comparison Test

Section 12-3
Two-Way ANOVA

Key Concepts

The analysis of variance procedure introduced in Section 12-2 is referred to as one-way analysis of variance because the data are categorized into groups according to a single factor (or treatment).

In this section we introduce the method of two-way analysis of variance, which is used with data partitioned into categories according to two factors.

Two-Way
Analysis of Variance

Two-Way ANOVA involves two factors.

The data are partitioned into subcategories called cells.

Definition

There is an interaction between two factors if the effect of one of the factors changes for different categories of the other factor.

Example:

Let's explore the data in Table 12-3 by calculating the mean for each cell and by constructing a graph.

Table 12-3 Chest Deceleration Measurements (in g) from Car Crash Tests

|	Size of Car		
	Small	**Medium**	**Large**
Foreign	44	41	32
	54	49	45
	43	47	42
Domestic	43	43	37
	44	37	38
	42	34	33

Example:

The individual cell means are shown in Table 12-4. Those means vary from a low of 36.0 to a high of 47.0, so they appear to vary considerably.

Table 12-4   Means (in g) of Cells from Table 12-3

	Size of Car		
	Small	**Medium**	**Large**
Foreign	47.0	45.7	39.7
Domestic	43.0	38.0	36.0

Example:

Figure 12-3 is an interaction graph, which shows graphs of those means, and that figure has two very notable features:
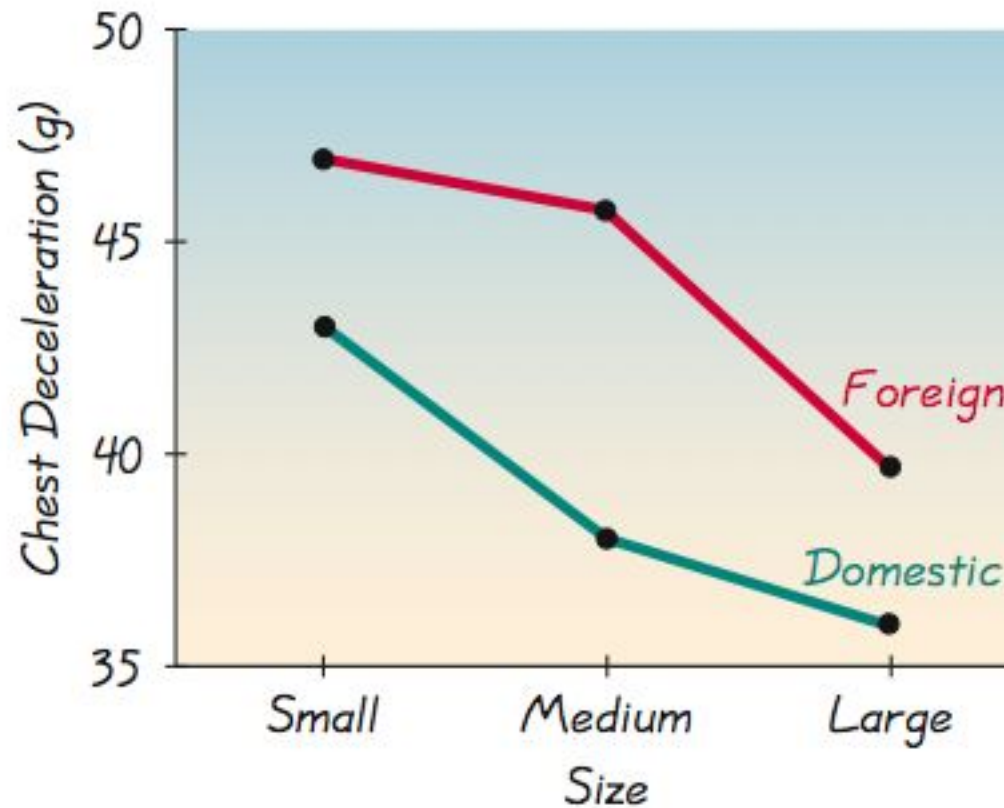
Example:

Larger means: Because the line segments representing foreign cars are higher than the line segments for domestic cars, it appears that foreign cars have consistently larger measures of chest deceleration.

Interaction: Because the line segments representing foreign cars appear to be approximately parallel to the line segments for domestic cars, it appears that foreign and domestic cars behave the same for the different car size categories, so there does not appear to be an interaction effect.
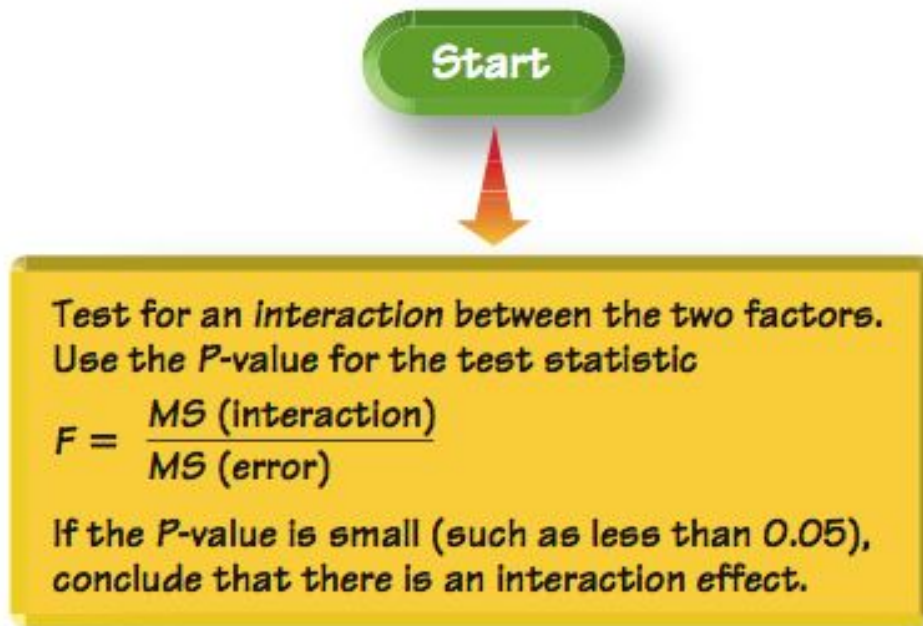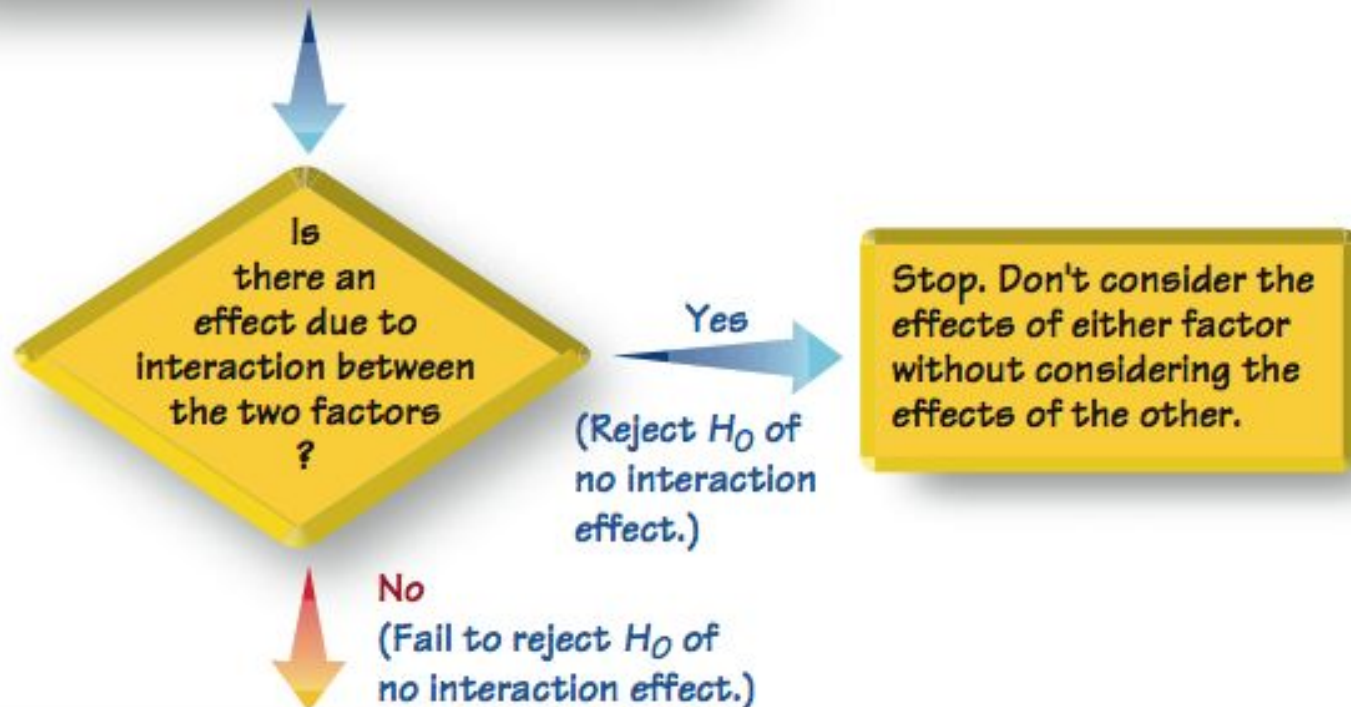
Example:

In general, if a graph such as Figure 12-3 results in line segments that are approximately parallel, we have evidence that there is not an interaction between the row and column variables.

These observations are largely subjective.

Let's proceed with a more objective method of two-way analysis of variance.

Figure 12- 4
Procedure for Two-Way ANOVAS

Test for an *interaction* between the two factors. Use the P-value for the test statistic

$$F = \frac{MS \text{ (interaction)}}{MS \text{ (error)}}$$

If the P-value is small (such as less than 0.05), conclude that there is an interaction effect.

Is there an effect due to interaction between the two factors ?

Yes
(Reject H_O of no interaction effect.)

Stop. Don't consider the effects of either factor without considering the effects of the other.

No
(Fail to reject H_O of no interaction effect.)

Test for effect from row factor using the P-value
for the test statistic

$$F = \frac{MS \text{ (row factor)}}{MS \text{ (error)}}$$

If the P-value is small (such as less than 0.05),
conclude that there is an effect from the
row factor.

Test for effect from column factor using the
P-value for the test statistic

$$F = \frac{MS \text{ (column factor)}}{MS \text{ (error)}}$$

If the P-value is small (such as less than 0.05),
conclude that there is an effect from the
column factor.

12.1 – *

Requirements

1. For each cell, the sample values come from a population with a distribution that is approximately normal.

2. The populations have the same variance σ^2.

3. The samples are simple random samples.

4. The samples are independent of each other.

5. The sample values are categorized two ways.

6. All of the cells have the same number of sample values.

Technology and Two-Way ANOVA

Two-Way ANOVA calculations are quite involved, so we will assume that a software package is being used.

Minitab, Excel, TI-83/4 or STATDISK can be used.

Procedure for Two-Way ANOVA

Step 1: Interaction Effect - test the null hypothesis that there is no interaction

Step 2: Row/Column Effects - if we conclude there is no interaction effect, proceed with these two hypothesis tests

Row Factor: no effects from row

Column Factor: no effects from column

Example:

Given the chest deceleration measurements in Table 12-3, use two-way analysis of variance to test for an interaction effect, an effect from the row factor of type of car (foreign, domestic), and an effect from the column factor of car size (small, medium, large). Use a 0.05 significance level.

Example:

Requirements are satisfied: sample values from normally distributed population; variances vary considerably but not extremely, we'll keep this in mind; simple random samples; independent samples, not matched; categorized two ways (foreign-domestic, small-medium-large; all cells has same number of samples (3)
Here's the Minitab results:

MINITAB

Source	DF	SS	MS	F	P
Type	1	117.556	117.556	5.44	0.038
Size	2	154.778	77.389	3.58	0.060
Interaction	2	14.778	7.389	0.34	0.717
Error	12	259.333	21.611		
Total	17	546.444			

Example:

Step 1: Interaction Effect

$$F = \frac{\text{MS(interaction)}}{\text{MS(error)}} = \frac{7.389}{21.611} = 0.34$$

P-value = 0.717, so fail to reject the null hypothesis of no interaction between the two factors

Step 2: Row/Column Effects

H_0: **There is no effect from the row factor (the row means are equal)**

H_0: **There is no effect from the column factor (the column means are equal)**

Example:

Row Factor:

$$F = \frac{\text{MS(type)}}{\text{MS(error)}} = \frac{117.556}{21.611} = 5.44$$

P-value = 0.038 < 0.05, so reject the null hypothesis of no effects from the type of car

Column Factor:

$$F = \frac{\text{MS(size)}}{\text{MS(error)}} = \frac{77.389}{21.611} = 3.58$$

P-value = 0.060 > 0.05, so fail to reject the null hypothesis of no effects from the size of car

Example:

Based on the sample data in Table 12-3, we conclude that chest deceleration measurements appear to be affected by whether the car is foreign or domestic, but those measurements do not appear to be affected by the size of the car.

Caution

Two-way analysis of variance is not one-way analysis of variance done twice. Be sure to test for an interaction between the two factors.

Special Case: One Observation per Cell and No Interaction

If our sample data consist of only one observation per cell, we lose *MS*(interaction), *SS*(interaction), and *df*(interaction).

If it seems reasonable to assume that there is no interaction between the two factors, make that assumption and then proceed as before to test the following two hypotheses separately:

H_0: There are no effects from the row factors.

H_0: There are no effects from the column factors.

Recap

In this section we have discussed:

❖ **Two- way analysis of variance (ANOVA)**

❖ **Special case: One observation per cell and no interaction**