Analysis of Variance

Chapter 12

Learning Objectives

LO12-1 Apply the F distribution to test a hypothesis that two population variances are equal

LO12-2 Use ANOVA to test a hypothesis that three or more population means are equal

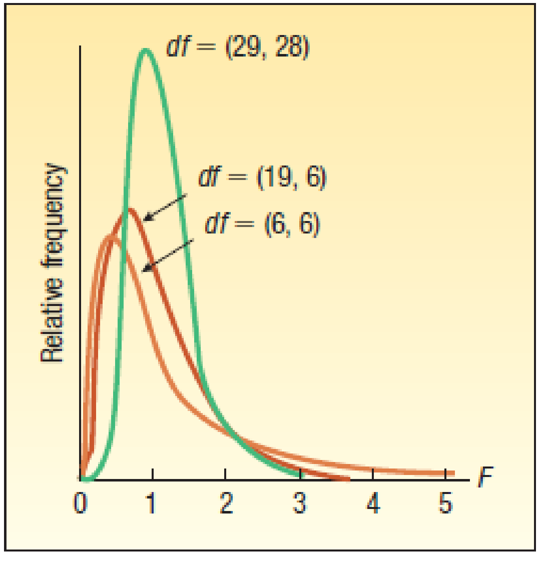
LO12-3 Use confidence intervals to test and interpret differences between pairs of population means

LO12-4 Use a blocking variable in a two-way ANOVA to test a hypothesis that three or more population means are equal

LO12-5 Perform a two-way ANOVA with interaction and describe the results

Characteristics of the F Distribution

- There is a family of F distributions. Each time the degrees of freedom in either the numerator or the denominator change, a new distribution is created
- The F distribution is continuous
- The F statistic cannot be negative
- The F distribution is positively skewed
- The F distribution is asymptotic



Comparing Two Population Variances

- The value of F is computed using the following equation

TEST STATISTIC FOR COMPARING TWO VARIANCES	$F = \frac{s_1^2}{s_2^2}$	(12–1)

- The larger of the two sample variances is placed in the numerator, forcing the ratio to be at least 1.00
- We calculate the standard deviation, s, and square the standard deviations to get the variance, s^2, for each population
- Example
- A health services corporation manages two hospitals in Knoxville: St. Mary's North and St. Mary's South. The mean waiting time in both Emergency Departments is 42 minutes. The hospital administrator believes St. Mary's North has more variation than St. Mary's South.

Compare Two Population Variances Example

Lammers Limos offers limousine service from Government Center in downtown Toledo, Ohio, to Metro Airport in Detroit. The president of the company is considering two routes. One is via U.S. 25 and the other via I-75. He wants to study the time it takes to get to the airport using each route and compare the results. He collected the following sample data. Using the .10 significance level, is there a difference in the variation in the driving times for the two routes?

U.S. Route 25	Interstate 75
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

U.S. ROUTE 25

$$\bar{x} = \frac{\Sigma x}{n} = \frac{408}{7} = 58.29 \qquad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{485.43}{7-1}} = 8.9947$$

INTERSTATE 75

$$\bar{x} = \frac{\Sigma x}{n} = \frac{472}{8} = 59.00 \qquad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{134}{8-1}} = 4.3753$$

Compare Two Population Variances Example Continued

Step 1: State the null and alternate hypothesis

$$H_0: \sigma^2_1 = \sigma^2_2$$
$$H_1: \sigma^2_1 \neq \sigma^2_2$$

Step 2: Select the level of significance, we decide to use .10

Step 3: Determine the test statistic, we'll use F

Step 4: State the decision rule, reject H_0 if the ratio of the sample variances > 3.87

Step 5: Compute the ratio of the two sample variances, it's 4.23 so we reject H_0

Step 6: We conclude there is a difference in the variation in the time to travel the two routes.

Degrees of Freedom for Denominator	Degrees of Freedom for Numerator			
	5	6	7	8
1	230	234	237	239
2	19.3	19.3	19.4	19.4
3	9.01	8.94	8.89	8.85
4	6.26	6.16	6.09	6.04
5	5.05	4.95	4.88	4.82
6	4.39	4.28	4.21	4.15
7	3.97	3.87	3.79	3.73
8	3.69	3.58	3.50	3.44
9	3.48	3.37	3.29	3.23
10	3.33	3.22	3.14	3.07

$$F = \frac{s_1^2}{s_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

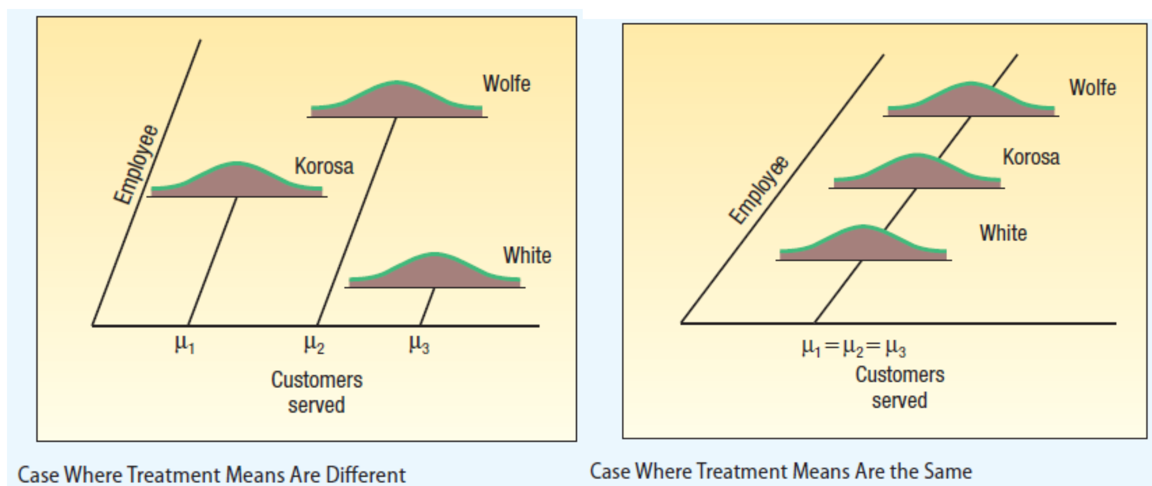
ANOVA: Analysis of Variance

- A one-way ANOVA is used to compare two or more treatment means
- ANOVA was first developed for use in agriculture; the term treatment was used to identify how different plots of land were treated with different fertilizers
- A treatment is a source of variation
- The assumptions underlying ANOVA are
  - The samples are from populations that follow the normal distribution
  - The populations have equal standard deviations
  - The populations are independent

ANOVA Example

Joyce Kuhlman manages a regional financial center. She wishes to compare the productivity, as measured by the number of customers served, among three employees. Four days are randomly selected and the number of customers served by each employee is recorded. Is there a difference in the mean number of customers served?

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48



The ANOVA Test

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48

First find the overall mean of the 12 observations. It is 58.
Next, find the difference between each particular value and the overall mean. Square these differences and sum up. This result is the total variation, here 1,082.

TOTAL VARIATION The sum of the squared differences between each observation and the overall mean.

Now, break this total variation in two components: variation due to treatment variation and random variation.

TREATMENT VARIATION The sum of the squared differences between each treatment mean and the grand or overall mean.

RANDOM VARIATION The sum of the squared differences between each observation and its treatment mean.

The ANOVA Test Continued

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48

Recall, the overall mean is 58 and the total variation is 1,082. Now, break this total variation in two components: variation due to treatment variation and random variation.

- The variation due to treatments is 992, found by squaring the difference between each treatment mean and the overall mean and then multiplying each squared difference by the number of observations in each treatment.

$$4(56-58)^2 + 4(70-58)^2 + 4(48-58)^2 = 992$$

- The random variation is 90, found by summing the squared differences between each value and the mean for each treatment.

$$(55-56)^2 + (54-56)^2 + \ldots + (48-48)^2 = 90$$

- Calculate the test statistic, F

$$F = \frac{992/2}{90/9} = 49.6$$

This ratio is quite different from 1, we can conclude there is a difference in the mean number of customers served by the three employees.

Finding the Value of F

- The formula for the sum of the squares total, SS total is

$$\text{SS total} = \Sigma(x - \bar{x}_G)^2 \quad [12\text{–}2]$$

where:
x is each sample observation.
$\bar{x}_G$ is the overall or grand mean.

- The formula for the sum of the squares error, SSE is

$$\text{SSE} = \Sigma(x - \bar{x}_c)^2 \quad [12\text{–}3]$$

where:
$\bar{x}_c$ is the sample mean for treatment c.

- The formula for the sum of the squares treatment, SST is

$$\text{SST} = \text{SS total} - \text{SSE} \quad [12\text{–}4]$$

- This information is summarized in the ANOVA table

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	SST	$k - 1$	$\text{SST}/(k - 1) = \text{MST}$	MST/MSE
Error	SSE	$n - k$	$\text{SSE}/(n - k) = \text{MSE}$	
Total	SS total	$n - 1$		

The case of the 4 airlines (ANOVA)

- Brunner randomly selected and surveyed passengers from the four airlines. Below is the sample information. Is there a difference in the mean satisfaction level among the four airlines? Use the .01 significance level.

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

The case of the 4 airlines (ANOVA)

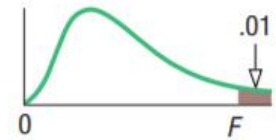
- Step 1: State the null and the alternate hypothesis
  - H_0: $\mu_N = \mu_W = \mu_P = \mu_B$
  - H_1: The mean scores are not all equal
- Step 2: Select the level of significance, we'll use .01
- Step 3: Determine the test statistic, the test statistic follows the F distribution
- Step 4: Formulate the decision rule (details on next slide)
- Step 5: Select the sample, calculate F , and make a decision,
- Step 6: Interpret the result, we conclude the populations are not all equal

The case of the 4 airlines (ANOVA) (step 4)

- To determine the decision rule, we need the critical value.
- The critical value for the *F* statistic for .01 significance level in appendix B.4.
- The degrees of freedom in the numerator and the denominator.
- The degrees of freedom in the numerator equals the number of treatments, designated as k, minus 1. ($k - 1 = 4 - 1 = 3$)
- The degrees of freedom in the denominator is the total number of observations, *n*, minus the number of treatments k. ($n - k = 22 - 4 = 18$)

The case of the 4 airlines (ANOVA) (Step 4)

- Refer to Appendix B.4 and the .01 significance level. Move horizontally across the top of the page to 3 degrees of freedom in the numerator.
- Then move down that column to the row with 18 degrees of freedom. The value at this intersection is 5.09. So the decision rule is to reject if the computed value of *F* exceeds 5.09.



Degrees of Freedom for the Denominator	Degrees of Freedom for the Num									
	1	2	3	4	5	6	7	8	9	10
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	605
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.8
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.6
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.8
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.2
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.8
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.5
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.3
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.1
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.9
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.8
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.6
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.5
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.5
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.4
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.3
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.3
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.2
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.2

The case of the 4 airlines (ANOVA) (Step 5)

- F=MST/MSE

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	SST	$k - 1$	$SST/(k - 1) = MST$	MST/MSE
Error	SSE	$n - k$	$SSE/(n - k) = MSE$	
Total	SS total	$n - 1$		

- You can determine these values by obtaining SS total and SSE, then finding SST by subtraction (remember this slide)

$$\bar{X}_G = \frac{1,664}{22} = 75.64$$

	Northern	WTA	Pocono	Branson	Total
	94	75	70	68	
	90	68	73	70	
	85	77	76	72	
	80	83	78	65	
		88	80	74	
			68	65	
			65		
Column total	349	391	510	414	1,664
n	4	5	7	6	22
Mean	87.25	78.20	72.86	69.00	75.64

The case of the 4 airlines (ANOVA) (Step 5, calculate SS Total)

- Deviation from the Grand Mean: $(X - \overline{X}_G) = 94 - 75.64 = 18.36.$

Northern	WTA	Pocono	Branson
18.36	−0.64	−5.64	−7.64
14.36	−7.64	−2.64	−5.64
9.36	1.36	0.36	−3.64
4.36	7.36	2.36	−10.64
	12.36	4.36	−1.64
		−7.64	−10.64
		−10.64	

- Square of the deviation $(X - \overline{X}_G)^2 = (94 - 75.64)^2 = (18.36)^2 = 337.09$

	Northern	WTA	Pocono	Branson	Total
	337.09	0.41	31.81	58.37	
	206.21	58.37	6.97	31.81	
	87.61	1.85	0.13	13.25	
	19.01	54.17	5.57	113.21	
		152.77	19.01	2.69	
			58.37	113.21	
			113.21		
Total	649.92	267.57	235.07	332.54	1,485.10

The case of the 4 airlines (ANOVA) (Step 5, calculate SSE)

- To compute the term SSE, find the deviation between each observation and its treatment mean. In the example, the mean of the first treatment (that is, the passengers on
- Northern Airlines) is 87.25, found by $\bar{X}_N = 349/4$. where N refers to Northern Airlines)
- Deviation from each treatment mean: $(X - \bar{X}_N) = (94 - 87.25) = 6.75$

Northern	WTA	Pocono	Branson
6.75	−3.2	−2.86	−1
2.75	−10.2	0.14	1
−2.25	−1.2	3.14	3
−7.25	4.8	5.14	−4
	9.8	7.14	5
		−4.86	−4
		−7.86	

The case of the 4 airlines (ANOVA) (Step 5, calculate SSE, SST, F Value)

- Square the deviation and sum the difference

	Northern	WTA	Pocono	Branson	Total
	45.5625	10.24	8.18	1	
	7.5625	104.04	0.02	1	
	5.0625	1.44	9.86	9	
	52.5625	23.04	26.42	16	
		96.04	50.98	25	
			23.62	16	
			61.78		
Total	110.7500	234.80	180.86	68	594.41

- SST: SST = SS total − SSE = 1,485.10 − 594.41 = 890.69.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	890.69	3	296.90	8.99
Error	594.41	18	33.02	
Total	1,485.10	21		

The case of the 4 airlines (ANOVA) (Step 6, decision)

- The computed value of *F* is 8.99, which is greater than the critical value of 5.09,so the null hypothesis is rejected. We conclude the population means are not all equal. The mean scores are not the same for the four airlines.
- It is likely that the passenger scores are related to the particular airline.
- At this point, we can only conclude there is a difference in the treatment means.
- We cannot determine which treatment groups differ or how many treatment groups differ.

Finding the Value of F Example

A group of four airlines hired Brunner Marketing Research Inc. to survey passengers regarding their level of satisfaction with a recent flight. Twenty-five questions offered a range of possible answers: excellent (4), good (3), fair (2), poor (1), so the highest possible score was 100. Brunner randomly selected and surveyed passengers from the four airlines. Is there a difference in the mean satisfaction level among the four airlines?

Step 1: State the null and the alternate hypothesis

$H_0: \mu_N = \mu_W = \mu_P = \mu_B$

H_1: The mean scores are not all equal

Step 2: Select the level of significance, we'll use .01

Step 3: Determine the test statistic, the test statistic follows the F distribution

Step 4: Formulate the decision rule, reject H_0 is F > 5.09

Step 5: Select the sample, calculate F (8.99), and make a decision, we reject H_0

Step 6: Interpret the result, we conclude the populations are not all equal

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	890.69	3	296.90	8.99
Error	594.41	18	33.02	
Total	1,485.10	21		

Pairs of Means

- If a null hypothesis of equal treatment means is rejected, we can identify the pairs of means that differ with the following confidence interval

CONFIDENCE INTERVAL FOR THE DIFFERENCE IN TREATMENT MEANS

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \textbf{(12–5)}$$

where:

$\bar{X}_1$ is the mean of the first sample.
$\bar{X}_2$ is the mean of the second sample.
t is obtained from Appendix B.2. The degrees of freedom is equal to $n - k$.
MSE is the mean square error term obtained from the ANOVA table [SSE/($n - k$)].
n_1 is the number of observations in the first sample.
n_2 is the number of observations in the second sample.

- If the confidence interval includes zero, there is not a difference between the treatment means

Pairs of Means Analysis Example

Recall in the previous example of airline satisfaction, we rejected the null hypothesis that the population means were equal; at least one of the airline's mean level of satisfaction is different from the others. But we do not know which pairs.

Use formula 12-5 to construct a confidence interval with the mean scores of Northern and Branson. Using a 95% level of confidence, we find the endpoints are 10.457 and 26.043.

Zero is not in the interval; so passengers on Northern rated service significantly different from those on Branson Airlines.

$$(\bar{x}_N - \bar{x}_B) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_N}+\frac{1}{n_B}\right)} = (87.25 - 69.00) \pm 2.101\sqrt{33.023\left(\frac{1}{4}+\frac{1}{6}\right)}$$
$$= 18.25 \pm 7.793$$

```
Branson subtracted from:

           Lower   Center   Upper   -----+---------+---------+---------+----
Pocono    -2.860    3.857  10.574               (----*---)
WTA        1.889    9.200  16.511                  (----*----)
Northern  10.457   18.250  26.043                        (----*----)
                                    -----+---------+---------+---------+----
                                       -15         0        15        30

Pocono subtracted from:

           Lower   Center   Upper   -----+---------+---------+---------+----
WTA       -1.726    5.343  12.412                (----*---)
Northern   6.826   14.393  21.960                      (----*----)
                                    -----+---------+---------+---------+----
                                       -15         0        15        30

WTA subtracted from:

           Lower   Center   Upper   -----+---------+---------+---------+----
Northern  0.951     9.050  17.149                 (----*----)
                                    -----+---------+---------+---------+----
                                       -15         0        15        30
```

A Two-Way ANOVA

- In a two-way ANOVA, we consider a second treatment variable
- This reduces the amount of error variance
- The second treatment variable is called the blocking variable
- It is determined using equation (12-6) below

$$SSB = k\Sigma(\bar{X}_b - \bar{X}_G)^2$$

- The SSE term, or sum of squares error, is found with the following equation

SUM OF SQUARES ERROR, TWO-WAY SSE = SS total − SST − SSB **(12–7)**

ANOVA Test Example

WARTA, the Warren Area Regional Transit Authority, is expanding bus service from the suburb of Starbrick to the business district of Warren. There are four routes being considered, U.S. 6, West End, Hickory St. , and Rte. 59. WARTA conducted tests to determine whether there is a difference in the mean travel times along the four routes; each driver drove each route. See the travel times in minutes for each driver-route combination below.

At the .05 significance level, is there a difference in the mean travel time along the four routes? If we remove the effects of the drivers, is there a difference in the mean travel time?

	Travel Time from Starbrick to Warren (minutes)			
Driver	U.S. 6	West End	Hickory St.	Rte. 59
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

ANOVA Test Example Continued

Step 1: State the null and alternate hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: Not all treatment means are the same

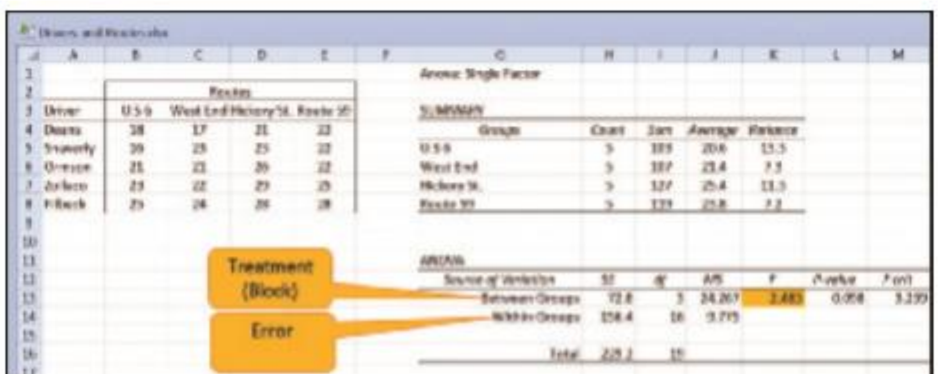
Step 2: Select the level of significance, we decide to use .05

Step 3: Select the test statistic, we use F

Step 4: State the decision rule, Reject H_0 if $F > 3.24$

Step 5: Make decision, F = 2.483, we do not reject the null hypothesis

Step 6: Interpret, there is no reason to conclude that any one of the routes is faster than any other.



The Blocking Variable

- In the WARTA example, we only considered the variation due to routes and took all other variables to be random
- Now, we'll include the variance due to the drivers by letting the drivers be the blocking variable

BLOCKING VARIABLE A second treatment variable that when included in the ANOVA analysis will have the effect of reducing the SSE term.

- To do so, requires that we calculate the SSB, the sum of the squares due to blocks

$$\text{SSB} = k\Sigma(\bar{x}_b - \bar{x}_G)^2$$

Two-Way Analysis of Variance

Including the variance of the drivers, here is a table of the drivers respective means with an overall mean of 22.8 minutes.

	Travel Time From Starbrick to Warren (minutes)					
Driver	U.S. 6	West End	Hickory St.	Rte. 59	Driver Sums	Driver Means
Deans	18	17	21	22	78	19.5
Snaverly	16	23	23	22	84	21
Ormson	21	21	26	22	90	22.5
Zollaco	23	22	29	25	99	24.75
Filbeck	25	24	28	28	105	26.25

Substituting this information in formula 12-6, we determine SSB is 119.7

$$SSB = k\Sigma(\bar{x}_b - \bar{x}_G)^2$$
$$= 4(19.5 - 22.8)^2 + 4(21.0 - 22.8)^2 + 4(22.5 - 22.8)^2 + 4(24.75 - 22.8)^2 + 4(26.25 - 22.8)^2$$
$$= 119.7$$

Then use formula 12-7 to find SSE

$$SSE = SS\ total - SST - SSB = 229.2 - 72.8 - 119.7$$
$$= 36.7$$

A Second Treatment Variable Continued

- Determine the F statistics for the treatment variable and the blocking variable from the following ANOVA table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	SST	$k - 1$	$SST/(k - 1) = MST$	MST/MSE
Blocks	SSB	$b - 1$	$SSB/(b - 1) = MSB$	MSB/MSE
Error	SSE	$(k - 1)(b - 1)$	$SSE/(k - 1)(b - 1) = MSE$	
Total	SS total	$n - 1$		

$$SSE = SS\ total - SST - SSB = 229.2 - 72.8 - 119.7 = 36.7$$

Source of Variation	(1) Sum of Squares	(2) Degrees of Freedom	(3) Mean Square (1)/(2)
Treatments	72.8	3	24.27
Blocks	119.7	4	29.93
Error	36.7	12	3.06
Total	229.2	19	

Hypothesis Test of Equal Block Means

Step 1: State the null hypothesis and the alternate hypotheses,

H_0: The treatment means are equal ($\mu_1 = \mu_2 = \mu_3 = \mu_4$)
H_1: At least one treatment mean is different

H_0: The block means are equal ($\mu_D = \mu_S = \mu_O = \mu_Z = \mu_F$)
H_1: At least one block mean is different

Step 2: Select the level of significance, we'll use .05
Step 3: Select the test statistic, we use F
Step 4: State the decision rule for the first set of hypotheses, reject H_0 if $F > 3.49$
Step 5: Make decision, the computed F ratio is 7.93 so we reject the null hypothesis that all treatment means are equal

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{24.27}{3.06} = 7.93$$

Step 6: Interpret, we conclude that at least one of the routes mean travel time is different from the other routes

Next, we test to find if the travel times for the various drivers are equal.

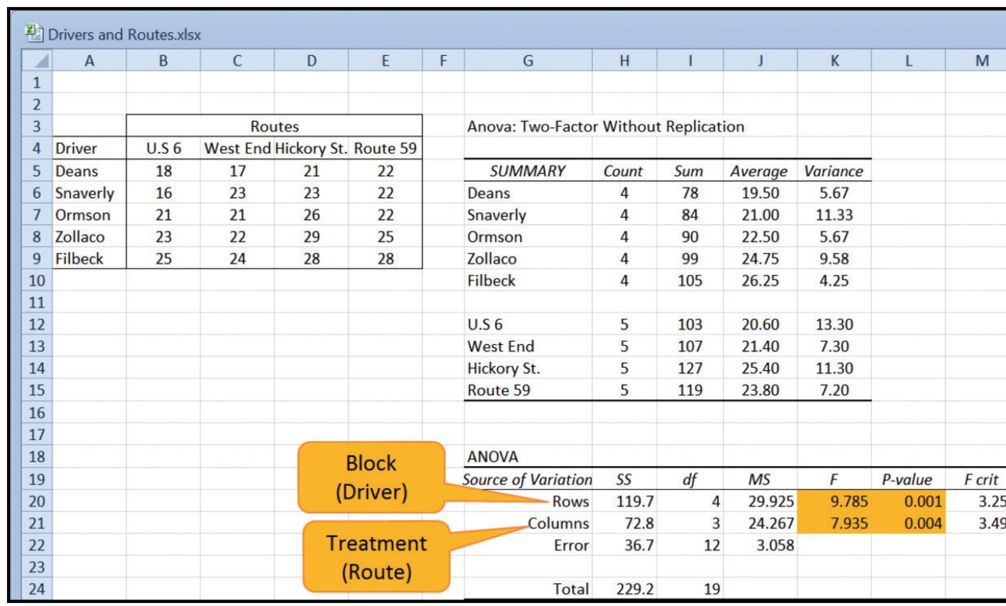
Hypothesis Test of Equal Block Means Continued

State the decision rule for the second set of hypotheses, reject H_0 if $F > 3.26$
Make a decision, the computed F ratio is 9.78 so we reject the null hypothesis

$$F = \frac{\text{MSB}}{\text{MSE}} = \frac{29.93}{3.06} = 9.78$$

Interpret, we conclude at least one driver's mean travel time is different from the others. WARTA management can conclude, based on the sample results, that there is a difference in the mean travel times of drivers.

Excel has a two-factor ANOVA procedure. The output for the WARTA example just completed is shown.



Driver	Routes: U.S 6	West End	Hickory St.	Route 59
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Deans	4	78	19.50	5.67
Snaverly	4	84	21.00	11.33
Ormson	4	90	22.50	5.67
Zollaco	4	99	24.75	9.58
Filbeck	4	105	26.25	4.25
U.S 6	5	103	20.60	13.30
West End	5	107	21.40	7.30
Hickory St.	5	127	25.40	11.30
Route 59	5	119	23.80	7.20

ANOVA

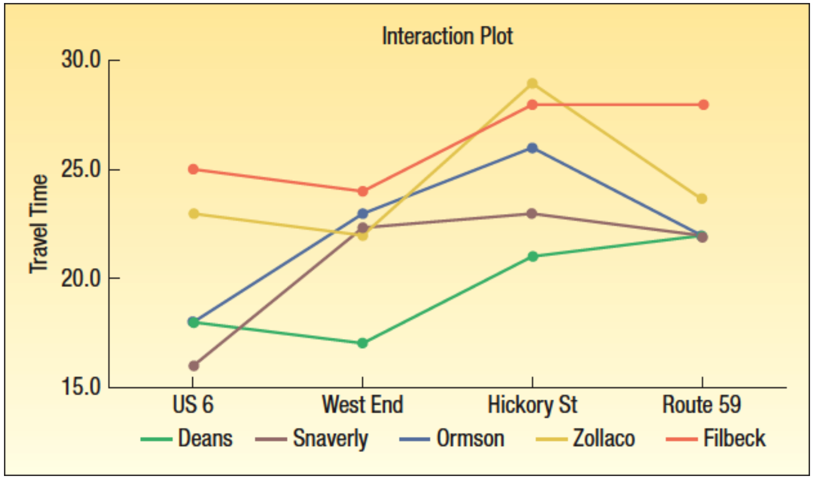
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	119.7	4	29.925	9.785	0.001	3.259
Columns	72.8	3	24.267	7.935	0.004	3.490
Error	36.7	12	3.058			
Total	229.2	19				

Interaction Plot

- An interaction plot illustrates the interaction of the two factors, route and driver
- Travel time is the response variable

INTERACTION The effect of one factor on a response variable differs depending on the value of another factor.

	Routes			
Drivers	U.S. 6	West End	Hickory	Rte. 59
Deans	18	17	21	22
Snaverly	16	22.33	23	22
Ormson	18	23	26	22
Zollaco	23	22	29	23.67
Filbeck	25	24	28	28



Hypothesis Tests for Interaction

- The next step is to investigate the interaction effects
  - Is there an interaction between drivers and routes?
  - Are the mean travel times for drivers the same?
  - Are the mean travel times for the routes the same?
- Test three sets of hypotheses
  - H_0: There is no interaction between drivers and routes
  - H_1: There is interaction between drivers and routes
  - H_0: The driver means are equal
  - H_1: At least one driver travel time mean is different
  - H_0: The route means are equal
  - H_1: At least one route travel time mean is different

ANOVA Table including Interactions

- The complete ANOVA table including interactions

Source of Variation	Sum of Squares	*df*	Mean Square	*F*
Factor A (driver)	SSA	$k - 1$	MSA = SSA/$(k - 1)$	MSA/MSE
Factor B (route)	SSB	$b - 1$	MSB = SSA/$(b - 1)$	MSB/MSE
Interaction	SSI	$(k - 1)(b - 1)$	MSI = SSI/$[(k - 1)(b - 1)]$	MSI/MSE
Error	SSE	$n - kb$	MSE = SSE/$(n - kb)$	
Total		$n - 1$		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Drivers	353.5667	4	88.39167	17.21916	0.0000	2.605975
Routes	244.9833	3	81.66111	15.90801	0.0000	2.838745
Interaction	125.7667	12	10.48056	2.041667	0.0456	2.003459
Error	205.3333	40	5.133333			
Total	929.65	59				

A One-Way ANOVA to Test a Hypothesis

- We will continue the analysis by conducting a one-way ANOVA for each route by testing the hypothesis

H_0: Driver times are equal

US 6; H_0: Driver times are equal

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	174	4	43.5	6.04167	0.010	3.478
Within Groups	72	10	7.2			
Total	246	14				

p-value < .05, reject the null

West End; H_0: Driver times are equal

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	88.6667	4	22.1667	4.05488	0.033	3.478
Within Groups	54.6667	10	5.46667			
Total	143.333	14				

p-value < .05, reject the null

Hickory; H_0: Driver times are equal

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	135.6	4	33.9	21.1875	0.000	3.478
Within Groups	16	10	1.6			
Total	151.6	14				

p-value < .05, reject the null

Route 59; H_0: Driver times are equal

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	81.0667	4	20.2667	3.23404	0.060	3.478
Within Groups	62.6667	10	6.26667			
Total	143.733	14				

p-value > .05, fail to reject the null

The results show there are significant differences in the mean travel times among the drivers for every route, except Route 59 which has a p-value of .06.