Mc Graw Hill

Because learning changes everything.®

Chapter 12

Analysis of Variance

Learning Objectives

LO12-1 Use ANOVA to test a hypothesis that three or more population means are equal.

LO12-2 Use confidence intervals to test and interpret differences between pairs of population means.

LO12-3 Use a blocking variable in a two-way ANOVA to test a hypothesis that three or more population means are equal.

ANOVA: Analysis of Variance [1]

انوفا = تستخدم للمقارنة بين 3 او اكثر لمعدل المجموعات (المعالجات)

A one-way ANOVA is used to compare three or more treatment means.

الافتراضات

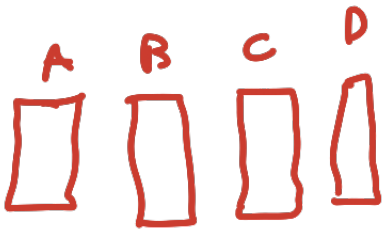
The assumptions underlying ANOVA are:

- The samples are from populations that follow normal distributions. يجب ان تتبع التوزيع الطبيعي
- The populations have equal standard deviations. المجتمعات لديها انحرافات معيارية متساوية
- The populations are independent. المجتمعات مستقله

When the conditions are met, the test statistic has an F-distribution. توزيعات F

Why do we need ANOVA, why not use the previous method based on the t-distribution?

ANOVA: Analysis of Variance 2



We could compare the population means two at a time.

Example: Suppose we have four groups.

- There are six comparisons we can make.
- Suppose we make each comparison at the 95% level.
- *P*(*All correct*) = $(.95)(.95)(.95)(.95)(.95)(.95) = .735$.
- The probability of at least one error is .265.
- The Type 1 error buildup is not satisfactory!

ANOVA allows us to compare all the means simultaneously.

Avoids the Type 1 error buildup.

ANOVA: Analysis of Variance

ANOVA was first developed for use in agriculture.

Example: Tomato plant heights for different amounts of fertilizer.

- Different plots of land would get different amounts of fertilizer.
- Treatment was used to identify the different plots/fertilizer amounts.

The term treatment identifies the different populations or groups.

ANOVA analyzes how the population means vary by the different treatments.

ANOVA: Analysis of Variance 4

- Example: Joyce Kuhlman manages a regional financial center.
- She wants 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?

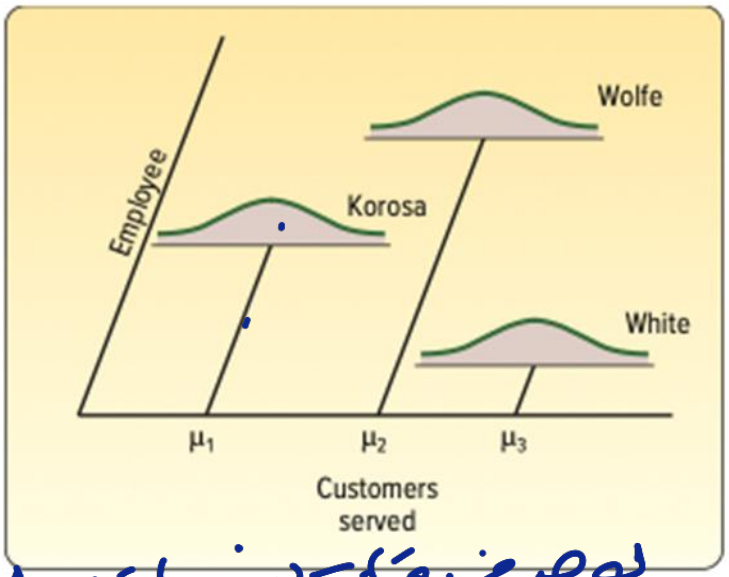
هل هناك فروقات بين عدد الزبائن المخدومين

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

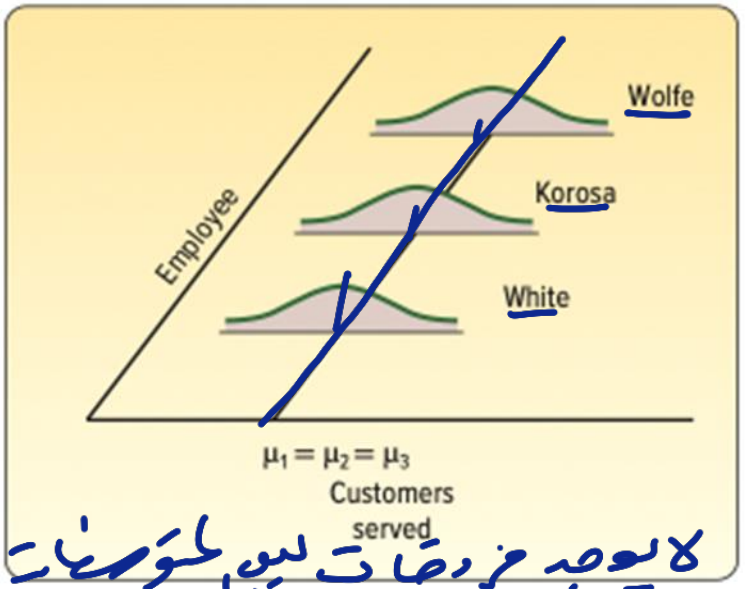


ANOVA: Analysis of Variance 5

- Example continued.
- Different vs Not Different.



يوجد فروقات بين المتوسطات



لا يوجد فروقات بين المتوسطات

Access the text alternative for these images.

ANOVA: Analysis of Variance 6

We want to determine if the population means are different.

Assume the population variances are the same.

ANOVA compares means through their variances.

Estimate the common population variances two ways.

The test statistic is the ratio of the two estimates.

- If the means are the same, the ratio should be 1.
- If the ratio is larger than 1, we might conclude the means are not the sample.

Use the F-distribution to determine when the ratio being greater than 1 occurred by chance.

ANOVA: Analysis of Variance

Total Variance = Treatment Variation + Random Variation

SS → SST → SSE

- Break the total variation into two parts.

Total Variation The sum of the squared differences between each observation and the overall mean.

مجموع مربعات الاختلافات بين كل البيانات و المتوسط الكلي

Treatment Variation The sum of the squared differences between each treatment mean and the grand or overall mean.

مجموع مربعات الفروقات بين كل متوسط لكل مجموعة مع المتوسط

Random Variation The sum of the squared differences between each observation and its treatment mean.

مجموع مربعات الفروقات بين كل قيمة و متوسط مجموعتها

ANOVA: Analysis of Variance 8

- The test statistic is the ratio of the estimates.
- $$F = \frac{\text{Between treatments estimate of the variances}}{\text{Within treatments estimate of the variances}} = \frac{MST}{MSE}$$
- Has an F-distribution.
- Numerator $df = k - 1$.
- Denominator $df = n - k$.
- k is the number of treatments.
- n is the total number of observations.

ANOVA: Analysis of Variance

- Total : $SS\ \text{total} = \sum(x - \bar{x}_G)^2$.
- Within : $SSE = \sum(x - \bar{x}_C)^2$.
- Between : *SST* = *SS total* − *SSE*.
- $\bar{x}_G$ is the overall grand mean.
- $\bar{x}_C$ is the sample mean for treatment *c*.
- Summarize with an ANOVA table.

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	S S T	$k - 1$	S S T/(k − 1) = M S T	M S T/M S E
Error	S S E	$n - k$	S S E/(n − k) = M S E	
Total	S S total	$n - 1$		

k: عدد المجموعات

	Wolfe	white	Korosa
	55	66	47
	54	76	51
	59	67	46
	56	71	48
	n=4	n=4	n=4
sum	224		
mean	56	70	48

① كتابة الفرضيات

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

$H_1: \mu_1 \neq \mu_2 = \mu_3$

② $\alpha = 0.05$

③ حساب قيمة $F_{Critical}$ من جدول F

$df_1 = k - 1 = 3 - 1 = 2$

$df_2 = n - k = 12 - 3 = 9$

$F_{critical} = 4.26$

X	$(X-XG)^2$	$(X-XC)^2$	
55	9	1	
54	16	4	
59	1	9	
56	4	0	
66	64	16	
76	324	36	
67	81	9	
71	169	1	
47	121	1	
51	49	9	
46	144	4	
48	100	0	
696	1082	90	992
58			
mean X G	SS TOTAL	SSE	SST

④ حساب المتوسطات لكل مجموعة $\bar{X}_C$

⑤ حساب grand mean $\bar{X}_G$

⑥ حساب SStotal

$SS\ total = \sum (X - \bar{X}_G)^2$

⑦ حساب SSE

$SSE = \sum (X - \bar{X}_C)^2$

⑧ حساب SST

$SST = SSTotal - SSE$

$= 1082 - 90 = 992$

⑨ نملأ جدول Anova بالبيانات المحسوبة

Anova

Source of Variation	SS	DF	MS	F	F-CRITICAL
Between Groups	SST 992	2	496	49.6	4.26
Within Groups	SSE 90	9	10		
Total	SS total 1082				

⑩ حساب MS

$MST = SST \div (k-1)$

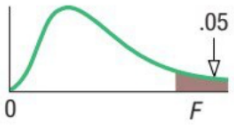
$MSE = SSE \div (n-k)$

$49.6 = \frac{496}{10} = \frac{MST}{MSE} =$ حساب F (11)

اتخاذ القرار (12) تتضح الفرضية $F > F_{cr}$

Appendix B

B.4 Critical Values of the *F* Distribution at a 5 Percent Level of Significance



Degrees of Freedom for the Denominator	Degrees of Freedom for the Numerator															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99

Sums of the Squares Treatment

The ***sum of the squares treatment***, denoted by SST, is calculates as

$$SST = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots \right) - \frac{(\sum x)^2}{n}$$

Where

- x = the score of a sample
- k = the number of different samples (or treatments)
- n_i = the size of sample i
- T_i = the sum of the values in sample i
- n_T = the number of values in all samples = $n_1 + n_2 + n_3 + \ldots$
- Σx = the sum of the values in all samples = $T_1 + T_2 + T_3 + \ldots$
- Σx^2 = the sum of the squares of the values in all samples

Total Sum of Squares

The ***Total sum of squares*** denoted by SSTotal, and is calculated as

$$SStotal = \sum x^2 - \frac{(\sum x)^2}{n}$$

Where

n = the number of values in all samples = $n_1 + n_2 + n_3 + \ldots$

Σx = the sum of the values in all samples = $T_1 + T_2 + T_3 + \ldots$

Σx^2 = the sum of the squares of the values in all samples

ANOVA: Analysis of Variance 10

- 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.

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	

- Is there a difference in the mean satisfaction level among the four airlines?

① $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1 = the mean are not equal

② $\alpha = 0.05$

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	

Data

data analysis

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Northern	4	349	87.25	36.91667
WTA	5	391	78.2	58.7
Pocono	7	510	72.85714	30.14286
Branson	6	414	69	13.6

$\alpha = 0.01$

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	890.6838	3	296.8946	8.990643	0.000743	3.159908
Within Groups	594.4071	18	33.02262			
Total	1485.091	21				

test statistic

F

$F_{critical}$

the test statistic is greater than critical value then reject null hypthiss

(4)

Northern	WTA	Pocono	Branson
94.00	75.00	70.00	68.00
90.00	68.00	73.00	70.00
85.00	77.00	76.00	72.00
80.00	83.00	78.00	65.00
	88.00	80.00	74.00
		68.00	65.00
$\bar{x}_G$	$\bar{x}_C$	65.00	
87.25	78.20	72.86	69.00

(3)

$df_1 = K - 1 = 4 - 1 = 3$

$df_2 = 22 - 4 = 18$

$F_{critical} = 3.16$

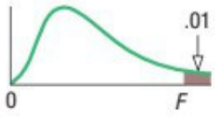
(5)

SS total SSE

X	$(X-XG)^2$	$(X-XC)^2$	
94.00	337.09	45.56	
90.00	206.21	7.56	
85.00	87.61	5.06	
80.00	19.01	52.56	
75.00	0.41	10.24	
68.00	58.37	104.04	
77.00	1.85	1.44	
83.00	54.17	23.04	
88.00	152.77	96.04	
70.00	31.81	8.18	
73.00	6.97	0.02	
76.00	0.13	9.86	
78.00	5.57	26.42	
80.00	19.01	50.98	
68.00	58.37	23.62	
65.00	113.21	61.78	
68.00	58.37	1.00	
70.00	31.81	1.00	
72.00	13.25	9.00	
65.00	113.21	16.00	
74.00	2.69	25.00	
65.00	113.21	16.00	
75.64	1485.09	594.41	890.7
mean X G	SS TOTAL	SSE	SST

Source of Variation	SS	DF	MS	F	F-CRITICAL
Between Groups	SST 890.7	3	296.9	8.99	3.16
Within Groups	SSE 594.41	18	33.02		
Total	SS 1485.09	21			

B.6B Critical Values of the *F* Distribution ($\alpha = .01$)



Degrees of Freedom for the Denominator	Degrees of Freedom for the Numerator															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59

ANOVA: Analysis of Variance 11

Example continued.

Step 1.

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$
- $H_1 :$ The mean scores are not all equal.

Step 2: Use 0.05

Step 3: $F = MTS/MSE.$

Step 4.

- $k = 4$, $n = 22$.
- Numerator $df = k - 1 = 4 - 1 = 3$.
- Denominator $df = n - k = 22 - 4 = 18$.
- The critical value is ~~3.09~~ 3.16.

ANOVA: Analysis of Variance 12

- Example continued.
- Step 5: $SS\text{total} = \sum (x - \bar{x}_G)^2 = 1{,}485.10.$

	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

Grand Mean

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	

	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

SS Total

Access the text alternative for these images.

ANOVA: Analysis of Variance 13

- Example continued.
- Step 5 continued: $SSE = \sum (x - \bar{x}_c)^2 = 594.41.$

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	

	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

SSE

- $SST = SS\text{total} - SSE = 1{,}485.10 - 594.41 = 890.69.$

Access the text alternative for these images.

ANOVA: Analysis of Variance 14

Example continued.

Step 5 continued.

- $MST = SST/k - 1 = 890.69/3 = 296.90.$
- $MSE = SSE/n - k = 594.41/18 = 33.02.$
- $F = MTS/MSE = 296.90/33.02 = 8.99.$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	890.69	3	296.90	8.99
Error	594.61	18	33.02	
Total	1,485.10	21		

- The test statistic is greater than the critical value, reject the null hypothesis.

3.16

Step 6: We can conclude the population means are not all equal.

ANOVA: Analysis of Variance 15

- Example continued.

Airline Anova

	A	B	C	D
1	Northern	WTA	Pocono	Branson
2	94	75	70	68
3	90	68	73	70
4	85	77	76	72
5	80	83	78	65
6		88	80	74
7			68	65
8			65	

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Northern	4	349	87.250	36.917
WTA	5	391	78.200	58.700
Pocono	7	510	72.857	30.143
Branson	6	414	69.000	13.600

0.05

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	890.684	3	296.895	8.99	0.0007	3.160
Within Groups	594.407	18	33.023			
Total	1485.091	21				

Access the text alternative for these images.

Inferences about Paris of Treatment Means

Suppose reject the ANOVA null hypothesis.

Conclude that all the treatment means are not the same.

We may want to know which treatment means differ.

There are several procedures available.

Use the confidence interval presented in Chapter 9.

- $$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{MSE * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

- Based on t with $df = n - k$.
- Assumes equal variances.
- *MSE* is the estimate of the common variance, like s_p^2.
- Confidence intervals that include 0 indicate no difference.

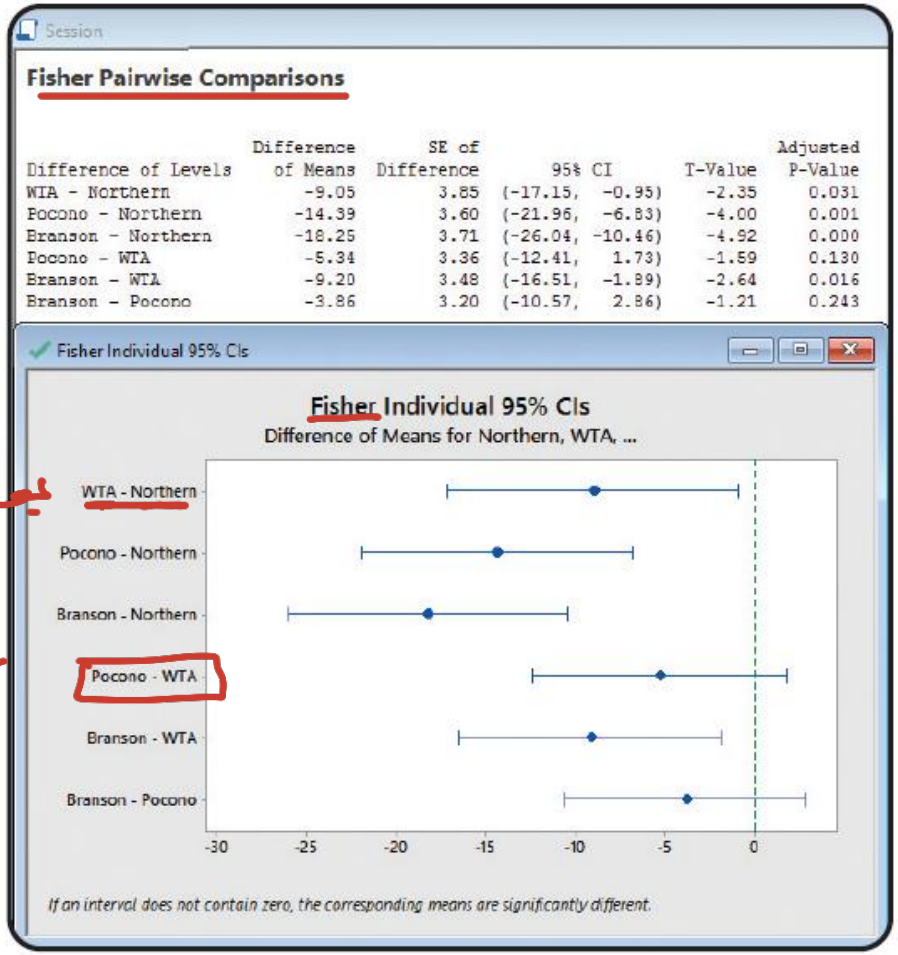
Inferences about Paris of Treatment Means 2

- Example: The previous airline example.

Session

Fisher Pairwise Comparisons

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
WIA - Northern	-9.05	3.85	(-17.15, -0.95)	-2.35	0.031
Pocono - Northern	-14.39	3.60	(-21.96, -6.83)	-4.00	0.001
Branson - Northern	-18.25	3.71	(-26.04, -10.46)	-4.92	0.000
Pocono - WTA	-5.34	3.36	(-12.41, 1.73)	-1.59	0.130
Branson - WTA	-9.20	3.48	(-16.51, -1.89)	-2.64	0.016
Branson - Pocono	-3.86	3.20	(-10.57, 2.86)	-1.21	0.243



يوجد اختلاف

و

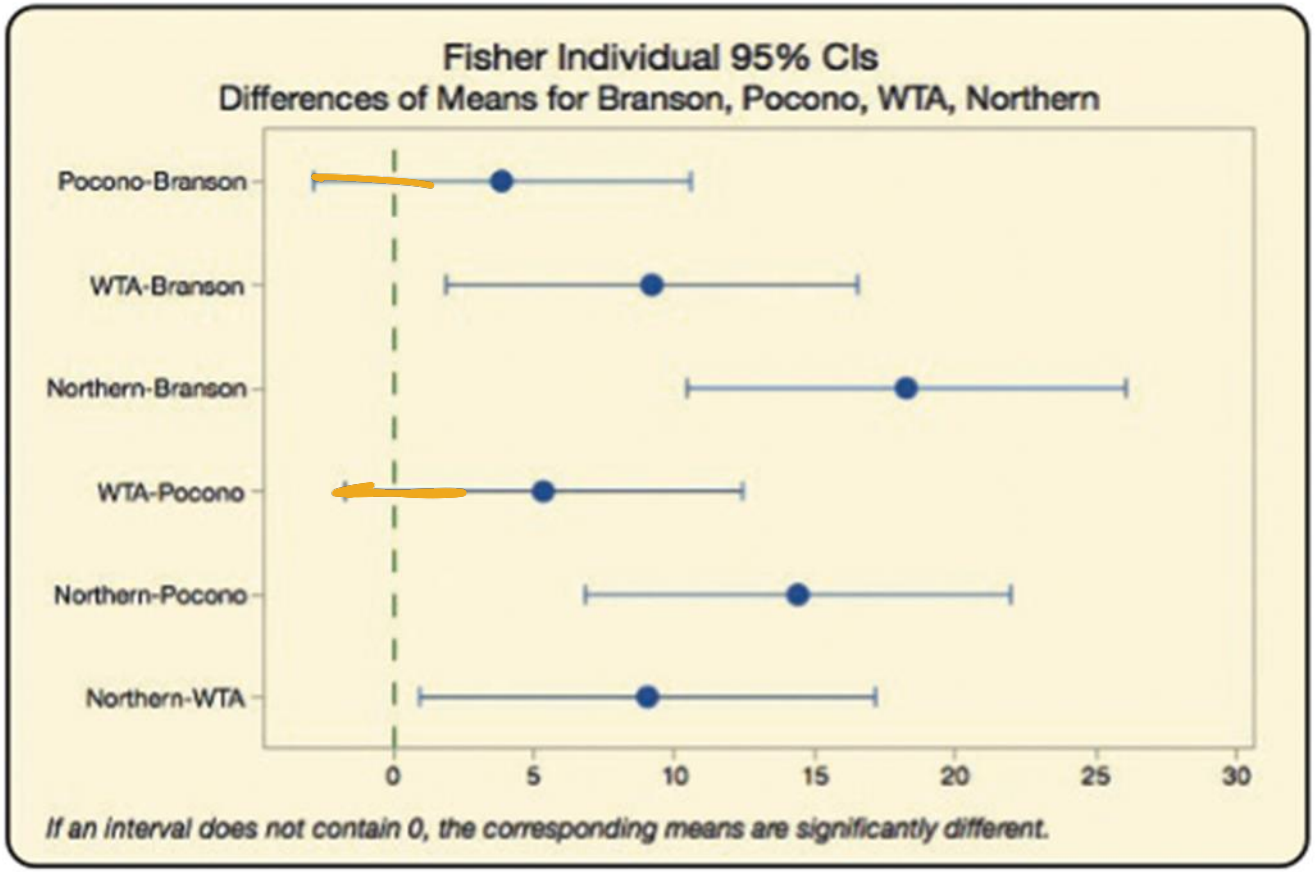
لا يوجد اختلاف

لا يوجد اختلافات

Access the text alternative for these images.

Inferences about Paris of Treatment Means 3

- Example continued.



Access the text alternative for these images.

Two-Way Analysis of Variance

We have only considered one source of variation.

Total = Between + Within

We called the within error the error/random variation.

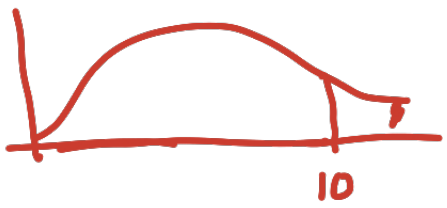
There might be other sources of variation to explain.

If we can explain more of the variation, then there is less error or random variation.

- Denominator of the F statistic will be smaller.
- Results in a larger F statistic.
- We might reject the null hypothesis.

Refer to other variables as blocking variables.

$$F = \frac{MST}{MSE}$$



one way anova

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

صبره	الرياضة	الرسام
25	41	50
30	24	49
40	15	35
15	33	25

$SSE = SS - SST$

SSE: Error / within group

SS: total

SST: Treatment- / Between groups

Two way anova test

	صبره	الرياضة	الرسام
عمر 15 سنوات	25	41	50
	30	24	49
	40	15	35
	15	33	25
عمر 11 سنه	20	15	40
	21	20	12
	23	35	18

يوجد متغير واحد المدة المدينه
متغير في الصفوف العمر

هل هناك اختلاف بين درجات الطلاب بسبب
العمر / المدينه

Two way test

Treatment = الاعمدة (المدن)

blocking = الصفوف (العمر)

$SSE = SS - SST - SSB$

	DF
SST	k-1
SSB	b-1
SSE	(k-1)(b-1)
SS	n-1

Two-Way Analysis of Variance 2

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

- $SSE = SS\ total - SST - SSB.$

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	$(k - 1)(b - 1)$	$SSE/(n - k) = MSE$	
Total	SS total	$n - 1$		

Two-Way Analysis of Variance 3

- Example: WARTA is expanding bus service from the suburb of Starbrick to the business district of Warren.
- There are four routes being considered and five drivers.

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

- 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?

① في حال اهمال تاثير المتغير هل هناك فرق في زمن الرحلة حسب سائق الباص

① كتابة الفرضيات

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

$H_1:$ Not all travels are the same

② تطبيق تحليل التوفا باستخدام الاكسل

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	5	103	20.6	13.3
Column 2	5	107	21.4	7.3
Column 3	5	127	25.4	11.3
Column 4	5	119	23.8	7.2

Test statistics

القيمة الحرجة

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	72.8	3	24.26666667	2.482523444	0.098105	3.238872
Within Groups	156.4	16	9.775			
Total	229.2	19				

③ القرار $2.48 < 3.238$

$F_{test\ statistic} < F_{critical\ value}$

Do not reject Null hypothysis

لايوجد ادله كافيه ان صنف اختلافات الاختلافات عشب ان تكون بسبب الصدفه

$0.098 > 0.05$ لا ترفض الفرضية الصفرية

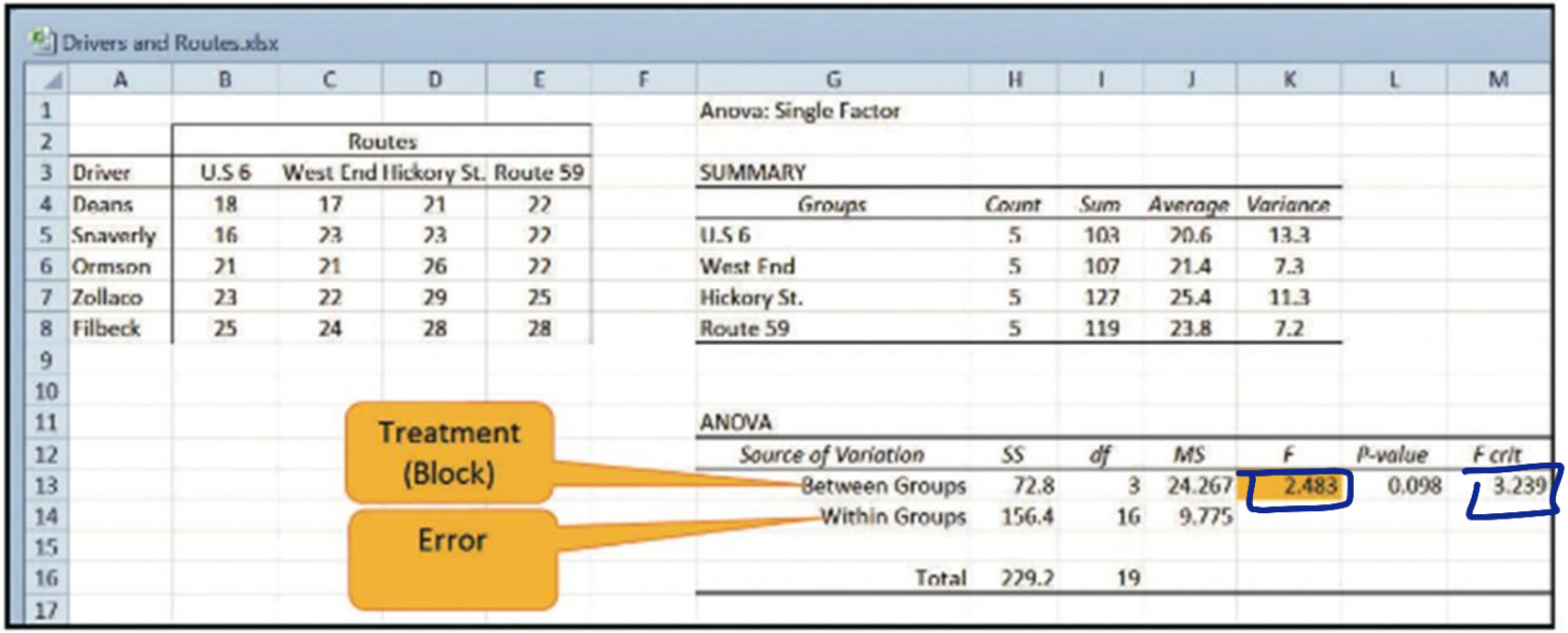
اذا كانت Pvalue اكبر من α (0.05) تقبل الفرضيه الصفريه

$p > \alpha$ do not reject H_0

$p < \alpha$ reject H_0

Two-Way Analysis of Variance 4

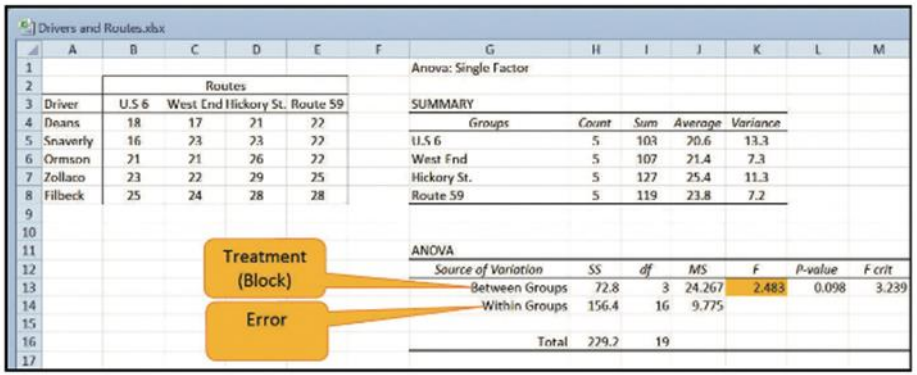
- Example continued.



Access the text alternative for these images.

Two-Way Analysis of Variance 5

- Example continued.



Drivers and Routes.xlsx

	A	B	C	D	E
1					
2		Routes			
3	Driver	U.S 6	West End	Hickory St.	Route 59
4	Deans	18	17	21	22
5	Snaverly	16	23	23	22
6	Ormson	21	21	26	22
7	Zollaco	23	22	29	25
8	Filbeck	25	24	28	28

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
U.S 6	5	103	20.6	13.3
West End	5	107	21.4	7.3
Hickory St.	5	127	25.4	11.3
Route 59	5	119	23.8	7.2

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	72.8	3	24.267	2.483	0.098	3.239
Within Groups	156.4	16	9.775			
Total	229.2	19				

Treatment (Block)

Error

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$
- H_1: Not all travel time means are the same.
- The p-value is .098, do not reject the null hypothesis.
- Differences in the mean travel time due to route could be due to chance.

الاختلافات سببها عشوائي ان تكون صدفة.

Access the text alternative for these images.

Two-Way Analysis of Variance 6

- Example continued.
- Let drivers be the blocking variable.

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

Driver	Travel Time from Starbrick to Warren (minutes) U.S. 6	Travel Time from Starbrick to Warren (minutes) West End	Travel Time from Starbrick to Warren (minutes) Hickory St.	Travel Time from Starbrick to Warren (minutes) Rte. 59	Travel Time from Starbrick to Warren (minutes) Driver Sums	Travel Time from Starbrick to Warren (minutes) Driver Means
Deans	18	17	21	22	78	19.50
Snavely	16	23	23	22	84	21.00
Ormson	21	21	26	22	90	22.50
Zollaco	23	22	29	25	99	24.75
Filbeck	25	24	28	28	105	26.25

blocking variable الاعتبار اخذ العوامل بعين

① كتابة الفرضيات

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

$H_1:$ Not all trave' time are same

② نطبق تحليل Two way Anova

SUMMARY	Count	Sum	Average	Variance			
Row 1	4	78	19.5	5.666666667			
Row 2	4	84	21	11.33333333			
Row 3	4	90	22.5	5.666666667			
Row 4	4	99	24.75	9.583333333			
Row 5	4	105	26.25	4.25			
Column 1	5	103	20.6	13.3			
Column 2	5	107	21.4	7.3			
Column 3	5	127	25.4	11.3			
Column 4	5	119	23.8	7.2			
ANOVA							
Source of Variation	SS	df	MS	F	P-value	F crit	
Rows	119.7	4	29.925	9.784741144	0.000934	3.259167	
Columns	72.8	3	24.26666667	7.934604905	0.003508	3.490295	
Error	36.7	12	3.058333333				
Total	229.2	19					

③ القرار (الصفوف)

نرفض الفرضية الصفرية $F_{statistic} > F_{critical}$

reject H_0

$9.78 > 3.25$

يوجد اختلافات في زمن الرحلة لمختلف السائق

$P_{value} < \alpha$ reject H_0

القرار (الاعمدة)

نرفض الفرضية الصفرية $7.9 > 3.4$ reject H_0

يوجد اختلافات في زمن الرحلة بسبب مسار الباص

Two-Way Analysis of Variance 7

- Example continued.

Drivers and Routes.xlsx

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				

Block (Driver)

Treatment (Route)

Access the text alternative for these images.

Two-Way Analysis of Variance 8

- Example continued.

Drivers and Routes.xlsx

	A	B	C	D	E
3		Routes			
4	Driver	U.S 6	West End	Hickory St.	Route 59
5	Deans	18	17	21	22
6	Snaverly	16	23	23	22
7	Ormson	21	21	26	22
8	Zollaco	23	22	29	25
9	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				

Block (Driver)

Treatment (Route)

الاعمدة

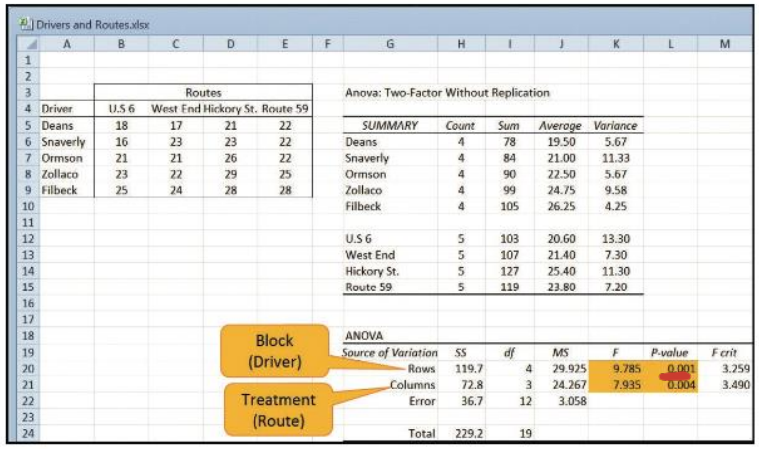
- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$
- H_1: Not all travel time means are the same.
- The p-value is 0.004, reject the null hypothesis.
- We can conclude there is a difference in the mean travel times for the different routes.

صنا له اختلاف في الرحلات حسب الاعمار

Access the text alternative for these images.

Two-Way Analysis of Variance 9

- Example continued.



Drivers and Routes.xlsx

Driver	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

Routes

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				

Block (Driver)

Treatment (Route)

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$.
- H_1: Not all travel time means are the same.
- The p-value is 0.001, reject the null hypothesis.
- We can conclude there is a difference in the mean travel times for the different drivers.

Access the text alternative for these images.

Chapter 12 Practice Problems

Question 1

Question 3

LO12-1

A real estate developer is considering investing in a shopping mall on the outskirts of Atlanta, Georgia. Three parcels of land are being evaluated. Of particular importance is the income in the area surrounding the proposed mall. A random sample of four families is selected near each proposed mall. Following are the sample results. At the .05 significance level, can the developer conclude there is a difference in the mean income?

a. What are the null and alternate hypotheses?

b. What is the critical value?

c. Compute the test statistic.

d. Compute the *p*-value.

e. What is your decision regarding the null hypothesis?

f. Interpret the result.

Southwyck Area ($000)	Franklin Park ($000)	Old Orchard ($000)
64	74	75
68	71	80
70	69	76
60	70	78

Data
Anova Single factor

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

H_1: not all sample have same mean

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	4	262	65.5	19.66667
Column 2	4	284	71	4.666667
Column 3	4	309	77.25	4.916667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	276.5	2	138.25	14.17949	0.001653	4.256495
Within Groups	87.75	9	9.75			
Total	364.25	11				

b) critical value = 4.256

c) test statistic = 14.179

d) pvalue = 0.0016

e)

Test Statistic > Critical value

14.179 > 4.256

reject H_0

f) the average income in each area are significantly different

Question 5

LO12-1, 2

The following are three observations collected from treatment 1, five observations collected from treatment 2, and four observations collected from treatment 3. Test the hypothesis that the treatment means are equal at the .05 significance level.

a. State the null hypothesis and the alternate hypothesis.

b. What is the decision rule?

c. Compute SST, SSE, and SS total.

d. Complete an ANOVA table.

e. Based on the value of the test statistic, state your decision regarding the null hypothesis.

f. If the null hypothesis is rejected, can we conclude that treatment 1 and treatment 2 differ? Use the 95% level of confidence.

Treatment 1	Treatment 2	Treatment 3
8	3	3
11	2	4
10	1	5
	3	4
	2	

$1 - 0.95$

$\alpha = 0.05$

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	3	29	9.666667	2.333333
Column 2	5	11	2.2	0.7
Column 3	4	16	4	0.666667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	107.2	2	53.6	50.95775	1.23E-05	4.256494729
Within Groups	9.466667	9	1.051852			
Total	116.6667	11				

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

H_1: Not all the means are equal

b) If test statistic (F) is > 4.26

reject H_0

c) $SST = 107.2$

$SSE = 9.47$

SS total $= 116.67$

d) Test statistic $>$ Critical value

$50.9 > 4.26$

Reject H_0

f) فترة الثقة

$$(\bar{x}_1 - \bar{x}_2) \mp t \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

t: من الجدول

$$(9.67 - 2.2) \pm 2.26 \sqrt{1.052\left(\frac{1}{3} + \frac{1}{5}\right)}$$

$$7.47 \pm 1.69$$

$$5.77 \qquad 9.16$$

$$[5.77 , 9.16]$$ وجهة افتراق

$Df = 9$ ← نذهب لـ t في الجدول

$\alpha = 0.05$ ←

$t = 2.26$

Question 7

Question 9

Question 11

LO12-3

Chapin Manufacturing Company operates 24 hours a day, 5 days a week. The workers rotate shifts each week. Management is interested in whether there is a difference in the number of units produced when the employees work on various shifts. A sample of five workers is selected and their output recorded on each shift. At the .05 significance level, can we conclude there is a difference in the mean production rate by shift or by employee?

Employee	Units Produced Day	Units Produced Afternoon	Units Produced Night
Skaff	31	25	35
Lum	33	26	33
Clark	28	24	30
Treece	30	29	28
Morgan	28	26	27

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Row 1	3	91	30.333	25.333
Row 2	3	92	30.667	16.333
Row 3	3	82	27.333	9.333
Row 4	3	87	29.000	1.000
Row 5	3	81	27.000	1.000
Column 1	5	150	30.000	4.500
Column 2	5	130	26.000	3.500
Column 3	5	153	30.600	11.300

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	33.733	4.000	8.433	1.552147239	0.276152	3.837853
Columns	62.533	2.000	31.267	5.754601227	0.028275	4.45897
Error	43.467	8.000	5.433			
Total	139.733	14.000				

for treatments (الاعمدة)

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

H_1: Not all means are equal

F > Critical value

5.75 > 4.46

Reject H_0

There is differens in production by the shifts

for blocks (الصفوف)

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1: Not all means are equal

$F < \text{Critical value}$

$1.55 < 3.84$

Do not reject H_0

There is no difference in production by employee

Because learning changes everything.®

www.mheducation.com