

Chapter 5

Excel 2007 Instructions	
One-Way Analysis of Variance 1. Open file: Digitron.xls. 2. On the Data tab, click Data Analysis . 3. Select ANOVA: Single Factor . 4. Define data range (columns B, C, D, E). 5. Specify alpha level 0.05. 6. Specify output location. 7. Click OK .	Randomized Complete Block Analysis of Variance 1. Open file: Citizens.xls. 2. On the Data tab, click Data Analysis . 3. Select ANOVA: Two-Factor Without Replication . 4. Define data range (include column A). 5. Specify alpha level = 0.05. 6. Indicate output location. 7. Click OK .

12.2. A manager is interested in testing whether three populations of interest have equal population means. Simple random samples of size 10 were selected from each population. The following ANOVA table and related statistics were computed:

ANOVA: Single Factor				
Summary				
Groups	Count	Sum	Average	Variance
Sample 1	10	507.18	50.72	35.06
Sample 2	10	405.79	40.58	30.08
Sample 3	10	487.64	48.76	23.13

ANOVA						
Source	SS	df	MS	F	p-value	F-crit
Between Groups	578.78	2	289.39	9.84	0.0006	3.354
Within Groups	794.36	27	29.42			
Total	1,373.14	29				

- a. State the appropriate null and alternative hypotheses.
- b. Conduct the appropriate test of the null hypothesis assuming that the populations have equal variances and the populations are normally distributed. Use $\alpha=0.05$ level of significance.
- c. If warranted, use the Tukey-Kramer procedure for multiple comparisons to determine which populations have different means. (Assume $\alpha = 0.05$.)

12-4. Respond to each of the following questions using this partially completed one-way ANOVA table:

Source of Variation	SS	df	MS	F-ratio
Between Samples	1,745			
Within Samples	_____	240		
Total	6,504	246		

- a. How many different populations are being considered in this analysis?
- b. Fill in the ANOVA table with the missing values.
- c. State the appropriate null and alternative hypotheses.

d. Based on the analysis of variance F -test, what conclusion should be reached regarding the null hypothesis? Test using a significance level of 0.01.

12-6. Given the following sample data

Item	Group 1	Group 2	Group 3	Group 4
1	20.9	28.2	17.8	21.2
2	27.2	26.2	15.9	23.9
3	26.6	21.6	18.4	19.5
4	22.1	29.7	20.2	17.4
5	25.3	30.3	14.1	
6	30.1	25.9		
7	23.8			

a. Based on the computations for the within- and between sample variation, develop the ANOVA table and test the appropriate null hypothesis using $\alpha = 0.05$. Use the p -value approach.

b. If warranted, use the Tukey-Kramer procedure to determine which populations have different means. Use $\alpha = 0.05$.

12-18. The following data were collected for a randomized block analysis of variance design with four populations and eight blocks:

	Group 1	Group 2	Group 3	Group 4
Block 1	56	44	57	84
Block 2	34	30	38	50
Block 3	50	41	48	52
Block 4	19	17	21	30
Block 5	33	30	35	38
Block 6	74	72	78	79
Block 7	33	24	27	33
Block 8	56	44	56	71

a. State the appropriate null and alternative hypotheses for the treatments and determine whether blocking is necessary.

b. Construct the appropriate ANOVA table.

c. Using a significance level equal to 0.05, can you conclude that blocking was necessary in this case? Use a test-statistic approach.

d. Based on the data and a significance level equal to 0.05, is there a difference in population means for the four groups? Use a p -value approach.

e. If you found that a difference exists in part d, use the LSD approach to determine which populations have different means.

12-19. A study was conducted to determine if differences in new textbook prices exist between on-campus bookstores, off-campus bookstores, and Internet bookstores. To control for differences in textbook prices that might exist across disciplines, the study randomly selected 12 textbooks and recorded the price of each of the 12 books at each of the three retailers. You may assume normality and equal variance assumptions have been met. The partially completed ANOVA table based on the study's findings is shown here:

ANOVA				
Source of Variation	SS	df	MS	F
Textbooks	16,624			
Retailer	2.4			
Error				
Total	17,477.6			

- Complete the ANOVA table by filling in the missing sums of squares, the degrees of freedom for each source, the mean square, and the calculated F -test statistic for each possible hypothesis test.
- Based on the study's findings, was it correct to block for differences in textbooks? Conduct the appropriate test at the $\alpha = 0.10$ level of significance.
- Based on the study's findings, can it be concluded that there is a difference in the average price of textbooks across the three retail outlets? Conduct the appropriate hypothesis test at the $\alpha = 0.10$ level of significance.

Answers

12.2

- The appropriate null and alternative hypotheses are:

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

$$H_A : \text{not all } \mu_j \text{ are equal}$$

- The one-way ANOVA test is appropriate for testing the null and alternative hypotheses. All the information needed is supplied in the table.
Using the F test approach, because $F = 9.84 > \text{critical } F = 3.35$, we reject the null hypothesis and conclude that the population means are not all equal.
Using the p-value approach, because $p\text{-value} = 0.0006 < \alpha = 0.05$, we reject the null hypothesis and conclude that the population means are not all equal.
- Because the null hypothesis has been rejected and we conclude that not all population means are equal, we can now apply the Tukey-Kramer method to determine which means are different. We start by calculating the Tukey-Kramer critical range value using:

$$\text{CriticalRange} = q_{1-\alpha} \sqrt{\frac{MSW}{2} \left\{ \frac{1}{n_i} + \frac{1}{n_j} \right\}}$$

The value for $q_{0.95}$ with $k = 3$ and $n_T - k = 30 - 3 = 27$ degrees of freedom is found in Appendix J as approximately 3.50. Because the sample sizes are equal in this situation, we need only compute one critical range value shown as follows:

$$\text{CriticalRange} = 3.5 \sqrt{\frac{29.42}{2} \left\{ \frac{1}{10} + \frac{1}{10} \right\}} = 6.0$$

We now compare all the possible contrasts of differences between sample means to the Tukey-Kramer critical range value.

Contrast	Significant ?
$ \bar{x}_1 - \bar{x}_2 = 50.72 - 40.58 = 10.14 > 6.0$	Yes
$ \bar{x}_1 - \bar{x}_3 = 50.72 - 48.76 = 1.96 < 6.0$	No
$ \bar{x}_2 - \bar{x}_3 = 40.58 - 48.76 = 8.18 > 6.0$	Yes

Thus, we conclude $\bar{x}_1 > \bar{x}_2$ and $\bar{x}_3 > \bar{x}_2$. Thus, the mean for population 2 is less than the means for the other two populations. However, the sample data do not provide sufficient evidence to conclude that the means for populations one and three are different.

12.4

a. $df_B + df_W = df_T \rightarrow df_B = df_T - df_W = 246 - 240 = 6 = k - 1 \rightarrow k = 7 = \text{number of populations.}$

b.

Source	SS	df	MS	F
Between Samples	1,745	6	290.833	14.667
Within Samples	4,759	240	19.829	
Total	6,504	246		

c. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$

H_A : At least two population means are different

d. F critical = 2.8778 (Minitab); from text table use $F_{6,200} = 2.893$

Since $14.667 > 2.8778$ reject H_0 and conclude that at least two populations means are different.

12.6

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Group 1	7	176	25.14286	10.00286
Group 2	6	161.9	26.98333	10.12567
Group 3	5	86.4	17.28	5.517
Group 4	4	82	20.5	7.553333

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	315.9324	3	105.3108	12.20025	0.000136	3.159911
Within Groups	155.3735	18	8.63186			
Total	471.3059	21				

a.

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

H_A : At least two population means are different

Since p-value = 0.000136 < 0.05 reject H_0 and conclude that at least two population means are different.

b. Tukey-Kramer Critical Range = $4.00 \sqrt{\frac{8.632}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$

Pair	Critical Range	Difference in Means	Significant?
1 vs 2	4.72	1.84	No
1 vs 3	5.137	7.863	Yes
1 vs 4	5.475	4.643	No
2 vs 3	4.968	9.703	Yes
2 vs 4	5.318	6.483	Yes
3 vs 4	5.691	3.22	No

12.18

- a. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
 H_A : At least two population means are different
 $H_0: \mu_{b1} = \mu_{b2} = \mu_{b3} = \mu_{b4} = \mu_{b5} = \mu_{b6} = \mu_{b7} = \mu_{b8}$
 H_A : Not all block means are equal

b.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Blocks	9123.375	7	1303.339	46.87669	2.08E-11	2.487582
Groups	1158.625	3	386.2083	13.8906	3.26E-05	3.072472
Error	583.875	21	27.80357			
Total	10865.88	31				

- c. Since $46.87669 > 2.487582$ reject H_0 and conclude that there is an indication that blocking was necessary.
d. Since p-value $0.0000326 < 0.05$ reject H_0 and conclude that at least two means are different.
e.

Least Significant
Difference (LSD) 5.48280844

	Mean Difference	Absolute Mean Difference	Significant
G1 v. G2	6.625	6.625	YES
G1 v. G3	-0.625	0.625	NO
G1 v. G4	-10.25	10.25	YES
G2 v. G3	-7.25	7.25	YES
G2 v. G4	-16.875	16.875	YES
G3 v. G4	-9.625	9.625	YES

12.19

- a. The calculations for the completed ANOVA table below are:
Textbooks (blocks) $df = b-1 = 12-1 = 11$
Retailer $df = k-1 = 3-1 = 2$
Error $df = (k-1)(b-1) = 11(2) = 22$
Total $df = n_t - 1$, where $n_t = (12 \text{ textbooks}) * (3 \text{ retailers}) = 36$
 $= 36 - 1 = 35$
SSW (error) = SST-SSBL-SSB = $17,477.6 - 16,624 - 2.4 = 851.2$
MSBL (Textbooks) = $16,624/11 = 1,511.3$
MSB (Retailer) = $2.4/2 = 1.2$
MSW (error) = $851.2/22 = 38.7$
F (textbooks) = $1,511.3/38.7 = 39.05$
F (Retailer) = $1.2/38.7 = 0.031$

ANOVA				
Source of Variation	SS	df	MS	F
Textbooks	16,624	11	1511.3	39.05
Retailer	2.4	2	1.2	0.031
Error	851.2	22	38.7	
Total	17477.6	35		

b.

$H_0: \mu_{On} = \mu_{Off} = \mu_I$

H_A : Not all populations have the same mean

Test to determine whether blocking is effective. Twelve textbooks were used to evaluate the prices at the three types of retail outlets. These constitute the blocks. The null and alternative hypotheses are:

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_{12}$

H_A : Not all block means are equal.

As shown in the ANOVA table from part (a), the F test statistic for this hypothesis test is the F for blocks (textbooks) = 39.05.

Using Excel's FINV function with $\alpha = 0.10$ and 11 and 22 degrees of freedom, $F_{\alpha, 11, 22} = 1.88$. Since $F = 39.05 > F_{\alpha, 11, 22} = 1.88$, reject the null hypothesis. This means that based on these sample data we can conclude that blocking is effective.

- c. We have three types of retail outlets (on-campus, off-campus, and Internet). The appropriate null and alternative hypotheses are:

$H_0: \mu_{On} = \mu_{Off} = \mu_I$

H_A : Not all populations have the same mean

As shown in the ANOVA table from part (a), the F test statistic for this null hypothesis is 0.031.

Using Excel's FINV function with $\alpha = 0.10$ and 2 and 22 degrees of freedom, $F_{\alpha, 2, 22} = 2.56$. Since $F = 0.031 < F_{\alpha, 2, 22} = 2.56$, do not reject the null hypothesis. Thus, based on these sample data we cannot conclude that there is a difference in textbook prices at the three different types of retail outlets.