


19. In an ANOVA table, MSE was equal to 10. Random samples of six were selected from each of four populations, where the sum of squares total was 250.
- Set up the null hypothesis and the alternate hypothesis.
 - What is the decision rule? Use the .05 significance level.
 - Complete the ANOVA table. What is the value of F ?
 - What is your decision regarding the null hypothesis?
20. The following is a partial ANOVA table.


Source	Sum of Squares	df	Mean Square	F
Treatment		2		
Error			20	
Total	500	11		

Complete the table and answer the following questions. Use the .05 significance level.

- How many treatments are there?
 - What is the total sample size?
 - What is the critical value of F ?
 - Write out the null and alternate hypotheses.
 - What is your conclusion regarding the null hypothesis?
21. A consumer organization wants to know whether there is a difference in the price of a particular toy at three different types of stores. The price of the toy was checked in a sample of five discount stores, five variety stores, and five department stores. The results are shown below. Use the .05 significance level. 

Discount	Variety	Department
\$12	\$15	\$19
13	17	17
14	14	16
12	18	20
15	17	19



22. Jacob Lee is a frequent traveler between Los Angeles and San Francisco. For the past month, he wrote down the flight times on three different airlines. The results are: 

Goust	Jet Red	Cloudtran
51	50	52
51	53	55
52	52	60
42	62	64
51	53	61
57	49	49
47	50	49
47	49	
50	58	
60	54	
54	51	
49	49	
48	49	
48	50	

- Use the .05 significance level and the five-step hypothesis-testing process to check if there is a difference in the mean flight times among the three airlines.
- Develop a 95% confidence interval for the difference in the means between Goust and Cloudtran.

Page
382-384
22.28

- a. Using analysis of variance techniques, test H_0 that the two mean test scores are equal; $\alpha = .05$.
- b. Using the t test from Chapter 11, compute t .
- c. Interpret the results.
26. There are four auto body shops in Bangor, Maine, and all claim to promptly serve customers. To check if there is any difference in service, customers are randomly selected from each repair shop and their waiting times in days are recorded. The output from a statistical software package is:

Summary				
Groups	Count	Sum	Average	Variance
Body Shop A	3	15.4	5.133333	0.323333
Body Shop B	4	32	8	1.433333
Body Shop C	5	25.2	5.04	0.748
Body Shop D	4	25.9	6.475	0.595833

ANOVA					
Source of Variation	SS	df	MS	F	p-value
Between Groups	23.37321	3	7.791069	9.612506	0.001632
Within Groups	9.726167	12	0.810514		
Total	33.09938	15			

Is there evidence to suggest a difference in the mean waiting times at the four body shops? Use the .05 significance level.

27. The fuel efficiencies for a sample of 27 compact, midsize, and large cars are entered into a statistical software package. Analysis of variance is used to investigate if there is a difference in the mean mileage of the three cars. What do you conclude? Use the .01 significance level.

Summary				
Groups	Count	Sum	Average	Variance
Compact	12	268.3	22.35833	9.388106
Midsize	9	172.4	19.15556	7.315278
Large	6	100.5	16.75	7.303

Additional results are shown below.

ANOVA					
Source of Variation	SS	df	MS	F	p-value
Between Groups	136.4803	2	68.24014	8.258752	0.001866
Within Groups	198.3064	24	8.262766		
Total	334.7867	26			



28. Three assembly lines are used to produce a certain component for an airliner. To examine the production rate, a random sample of six hourly periods is chosen for each assembly line and the number of components produced during these periods for each line is recorded. The output from a statistical software package is:

Summary				
Groups	Count	Sum	Average	Variance
Line A	6	250	41.66667	0.266667
Line B	6	260	43.33333	0.666667
Line C	6	249	41.5	0.7

Continued on Next Page

ANOVA					
Source of Variation	SS	df	MS	F	p-value
Between Groups	12.33333	2	6.166667	11.32653	0.001005
Within Groups	8.166667	15	0.544444		
Total	20.5	17			

- a. Use a .01 level of significance to test if there is a difference in the mean production of the three assembly lines.
- b. Develop a 99% confidence interval for the difference in the means between Line B and Line C.
29. The postal service groups first-class mail as letters, cards, flats, or parcels. Over a period of three weeks, one item of each kind was sent from a particular postal administrative center. The total time in transit was recorded. A statistical software package was then used to perform the analysis. The results follow.

Source	DF	SS	MS	F	P
Factor	3	13.82	4.61	2.72	0.051
Error	68	115.17	1.69		
Total	71	128.99			

S = 1.301 R-Sq = 10.71% R-Sq(adj) = 6.77%

Individual 95% CIs for Mean Based on Pooled StDev

Level	N	Mean	StDev	
Letters	18	1.444	1.097	(-----*-----)
Cards	18	1.667	1.455	(-----*-----)
Flats	18	2.444	1.617	(-----*-----)
Parcels	18	2.389	0.916	(-----*-----)

1.20 1.80 2.40 3.00

Use the .05 significance level to test if this evidence suggests a difference in the means for the different types of first-class mail.

30. For your email, you use a filter to block spam from your inbox. The number of items blocked by day of week is recorded and a statistical software system is used to perform the analysis that follows. Here are the results:

Source	DF	SS	MS	F	P
Factor	6	1367.8	228.0	5.72	0.000
Error	48	1913.2	39.9		
Total	54	3281.0			

S = 6.313 R-Sq = 41.69% R-Sq(adj) = 34.40%

Individual 95% CIs for Mean Based on Pooled StDev

Level	N	Mean	StDev	
Monday	10	74.000	6.164	(-----*-----)
Tuesday	9	66.111	7.288	(-----*-----)
Wednesday	7	74.143	2.268	(-----*-----)
Thursday	8	62.375	5.041	(-----*-----)
Friday	8	75.125	4.454	(-----*-----)
Saturday	5	63.200	7.259	(-----*-----)
Sunday	8	72.375	9.164	(-----*-----)

60.0 66.0 72.0 78.0

Use the .05 significance level to test if this evidence suggests a difference in the means for the different days of the week.

31. Investors can now make stock trades online for as little as \$7 per trade. Some suggest that this will motivate investors to increase the percentage of stocks in their

Salary (\$ thousands), Y	Years of Experience, X_1	Principal's Rating, X_2	Master's Degree,* X_3
31.1	8	35	0
33.6	5	43	0
29.3	2	51	1
\vdots	\vdots	\vdots	\vdots
30.7	4	62	0
32.8	2	80	1
42.8	8	72	0

*1 = yes, 0 = no.


- Develop a correlation matrix. Which independent variable has the strongest correlation with the dependent variable? Does it appear there will be any problems with multicollinearity?
- Determine the regression equation. What salary would you estimate for a teacher with five years' experience, a rating by the principal of 60, and no master's degree?
- Conduct a global test of hypothesis to determine whether any of the regression coefficients differ from zero. Use the .05 significance level.
- Conduct a test of hypothesis for the individual regression coefficients. Would you consider deleting any of the independent variables? Use the .05 significance level.
- If your conclusion in part (d) was to delete one or more independent variables, run the analysis again without those variables.
- Determine the residuals for the equation of part (e). Use a histogram to verify that the distribution of the residuals is approximately normal.
- Plot the residuals computed in part (f) in a scatter diagram with the residuals on the Y-axis and the \hat{Y} values on the X-axis. Does the plot reveal any violations of the assumptions of regression?

19. A consumer analyst collected the following data on the screen sizes of popular LCD televisions sold recently at a large retailer:

Manufacturer	Screen	Price
Sharp	46	\$1473.00
Samsung	52	2300.00
Samsung	46	1790.00
Sony	40	1250.00
Sharp	42	1546.50
Samsung	46	1922.50
Samsung	40	1372.00
Sharp	37	1149.50
Sharp	46	2000.00
Sony	40	1444.50
Sony	52	2615.00
Samsung	32	747.50
Sharp	37	1314.50
Sharp	32	853.50
Sharp	52	2778.00
Samsung	40	1749.50
Sharp	32	1035.00
Samsung	52	2950.00
Sony	40	1908.50
Sony	52	3103.00
Sony	46	2606.00
Sony	46	2861.00
Sony	52	3434.00


Continued On next Page

Page
488-494
#19,30

- a. Does there appear to be a linear relationship between the screen size and the price?
 - b. Which variable is the "dependent" variable?
 - c. Using statistical software, determine the regression equation. Interpret the value of the slope in the regression equation.
 - d. Include the manufacturer in a multiple linear regression analysis using a "dummy" variable. Does it appear that some manufacturers can command a premium price?
Hint: You will need to use a set of indicator variables.
 - e. Test each of the individual coefficients to see if they are significant.
 - f. Make a plot of the residuals and comment on whether they appear to follow a normal distribution.
 - g. Plot the residuals versus the fitted values. Do they seem to have the same amount of variation?
20. A regional planner is studying the demographics in a region of a particular state. She has gathered the following data on nine counties. 

County	Median Income	Median Age	Coastal
A	\$48,157	57.7	1
B	48,568	60.7	1
C	46,816	47.9	1
D	34,876	38.4	0
E	35,478	42.8	0
F	34,465	35.4	0
G	35,026	39.5	0
H	38,599	65.6	0
J	33,315	27.0	0

- a. Is there a linear relationship between the median income and median age?
 - b. Which variable is the "dependent" variable?
 - c. Use statistical software to determine the regression equation. Interpret the value of the slope in a simple regression equation.
 - d. Include the aspect that the county is "coastal" or not in a multiple linear regression analysis using a "dummy" variable. Does it appear to be a significant influence on incomes?
 - e. Test each of the individual coefficients to see if they are significant.
 - f. Make a plot of the residuals and comment on whether they appear to follow a normal distribution.
 - g. Plot the residuals versus the fitted values. Do they seem to have the same amount of variation?
21. Great Plains Roofing and Siding Company Inc. sells roofing and siding products to home repair retailers, such as Lowe's and Home Depot, and commercial contractors. The owner is interested in studying the effects of several variables on the value of shingles sold (\$000). The marketing manager is arguing that the company should spend more money on advertising, while a market researcher suggests it should focus more on making its brand and product more distinct from its competitors.

The company has divided the United States into 26 marketing districts. In each district, it collected information on the following variables: volume of sales (in thousands of dollars), advertising dollars (in thousands), number of active accounts, number of competing brands, and a rating of district potential. 

Sales (000s)	Advertising Dollars (000s)	Number of Accounts	Number of Competitors	Market Potential
79.3	5.5	31	10	8
200.1	2.5	55	8	6
163.2	8.0	67	12	9

(continued)



- f. Rerun the analysis until only significant regression coefficients remain in the analysis. Identify these variables.
 - g. Develop a histogram of the residuals from the final regression equation developed in part (f). Is it reasonable to conclude that the normality assumption has been met?
 - h. Plot the residuals against the fitted values from the final regression equation developed in part (f). Plot the residuals on the vertical axis and the fitted values on the horizontal axis.
30. Refer to the Baseball 2010 data, which report information on the 30 Major League Baseball teams for the 2010 season. Let the number of games won be the dependent variable and the following variables be independent variables: team batting average, number of stolen bases, number of errors committed, team ERA, number of home runs, and whether the team plays in the American or the National League. Add a league code variable using 0 for the National League and 1 for the American League.
- a. Use a statistical software package to determine the multiple regression equation. Discuss each of the variables. For example, are you surprised that the regression coefficient for ERA is negative? Is the number of wins affected by whether the team plays in the National or the American League?
 - b. Find the coefficient of determination for this set of independent variables.
 - c. Develop a correlation matrix. Which independent variables have strong or weak correlations with the dependent variable? Do you see any problems with multicollinearity?
 - d. Conduct a global test on the set of independent variables. Interpret.
 - e. Conduct a test of hypothesis on each of the independent variables. Would you consider deleting any of the variables? If so, which ones?
 - f. Rerun the analysis until only significant net regression coefficients remain in the analysis. Identify these variables.
 - g. Develop a histogram of the residuals from the final regression equation developed in part (f). Is it reasonable to conclude that the normality assumption has been met?
 - h. Plot the residuals against the fitted values from the final regression equation developed in part (f). Plot the residuals on the vertical axis and the fitted values on the horizontal axis.
31. Refer to the Buena School District bus data. First, add a variable to change the type of bus (diesel or gasoline) to a qualitative variable. If the bus type is diesel, then set the qualitative variable to 0. If the bus type is gasoline, then set the qualitative variable to 1. Develop a regression equation using statistical software with maintenance as the dependent variable and age, miles, and bus type as the independent variables.
- a. Write out the multiple regression equation analysis. Discuss each of the variables.
 - b. Determine the value of R^2 . Interpret.
 - c. Develop a correlation matrix. Which independent variables have strong or weak correlations with the dependent variable? Do you see any problems with multicollinearity?
 - d. Conduct the global test on the set of independent variables. Interpret.
 - e. Conduct a test of hypothesis on each of the independent variables. Would you consider deleting any of the variables? If so, which ones?
 - f. Rerun the analysis until only significant regression coefficients remain in the analysis. Identify these variables.
 - g. Develop a histogram of the residuals from the final regression equation developed in part (f). Is it reasonable to conclude that the normality assumption has been met?
 - h. Plot the residuals against the fitted values from the final regression equation developed in part (f) against the fitted values of \hat{Y} . Plot the residuals on the vertical axis and the fitted values on the horizontal axis.

Practice Test

Part 1—Objective

- 1. Multiple regression analysis describes the relationship between one dependent variable and two or more _____.
- 2. In multiple regression analysis, the regression coefficients are computed using the method of _____. (residuals, normality, least squares, standardization)
- 3. In multiple regression analysis, the multiple standard error of the estimate is the square root of the _____. (mean square error, residual, residual squared, explained variation)