

- b. Let days-on-the-market be the dependent variable and price be the independent variable. Determine the regression equation. Estimate the days-on-the-market of a home that is priced at \$300,000. Determine the 95% confidence interval of days-on-the-market for homes with a mean price of \$300,000, and the 95% prediction interval of days-on-the-market for a home priced at \$300,000.
 - c. Can you conclude that the independent variables "days on the market" and "selling price" are positively correlated? Are the size of the home and the selling price positively correlated? Use the .05 significance level. Report the p -value of the test. Summarize your results in a brief report.
63. **FILE** Refer to the Baseball 2016 data, which reports information on the 2016 Major League Baseball season. Let attendance be the dependent variable and total team salary be the independent variable. Determine the regression equation and answer the following questions.
- a. Draw a scatter diagram. From the diagram, does there seem to be a direct relationship between the two variables?
 - b. What is the expected attendance for a team with a salary of \$100.0 million?
 - c. If the owners pay an additional \$30 million, how many more people could they expect to attend?
 - d. At the .05 significance level, can we conclude that the slope of the regression line is positive? Conduct the appropriate test of hypothesis.
 - e. What percentage of the variation in attendance is accounted for by salary?
 - f. Determine the correlation between attendance and team batting average and between attendance and team ERA. Which is stronger? Conduct an appropriate test of hypothesis for each set of variables.
64. **FILE** Refer to the Lincolnville School bus data. Develop a regression equation that expresses the relationship between age of the bus and maintenance cost. The age of the bus is the independent variable.
- a. Draw a scatter diagram. What does this diagram suggest as to the relationship between the two variables? Is it direct or indirect? Does it appear to be strong or weak?
 - b. Develop a regression equation. How much does an additional year add to the maintenance cost. What is the estimated maintenance cost for a 10-year-old bus?
 - c. Conduct a test of hypothesis to determine whether the slope of the regression line is greater than zero. Use the .05 significance level. Interpret your findings from parts (a), (b), and (c) in a brief report.

- b. Let days-on-the-market be the dependent variable and price be the independent variable. Determine the regression equation. Estimate the days-on-the-market of a home that is priced at \$300,000. Determine the 95% confidence interval of days-on-the-market for homes with a mean price of \$300,000, and the 95% prediction interval of days-on-the-market for a home priced at \$300,000.
- c. Can you conclude that the independent variables "days on the market" and "selling price" are positively correlated? Are the size of the home and the selling price positively correlated? Use the .05 significance level. Report the p -value of the test. Summarize your results in a brief report.
63. **FILE** Refer to the Baseball 2016 data, which reports information on the 2016 Major League Baseball season. Let attendance be the dependent variable and total team salary be the independent variable. Determine the regression equation and answer the following questions.
- Draw a scatter diagram. From the diagram, does there seem to be a direct relationship between the two variables?
 - What is the expected attendance for a team with a salary of \$100.0 million?
 - If the owners pay an additional \$30 million, how many more people could they expect to attend?
 - At the .05 significance level, can we conclude that the slope of the regression line is positive? Conduct the appropriate test of hypothesis.
 - What percentage of the variation in attendance is accounted for by salary?
 - Determine the correlation between attendance and team batting average and between attendance and team ERA. Which is stronger? Conduct an appropriate test of hypothesis for each set of variables.
64. **FILE** Refer to the Lincolnville School bus data. Develop a regression equation that expresses the relationship between age of the bus and maintenance cost. The age of the bus is the independent variable.
- Draw a scatter diagram. What does this diagram suggest as to the relationship between the two variables? Is it direct or indirect? Does it appear to be strong or weak?
 - Develop a regression equation. How much does an additional year add to the maintenance cost. What is the estimated maintenance cost for a 10-year-old bus?
 - Conduct a test of hypothesis to determine whether the slope of the regression line is greater than zero. Use the .05 significance level. Interpret your findings from parts (a), (b), and (c) in a brief report.

- d. Conduct an individual test to determine whether any of the independent variables can be dropped.
- e. Rerun the regression equation, using only the independent variables that are significant. How much more does a man earn per month than a woman? Does it make a difference whether the employee has a management or engineering position?
26. **FILE** Many regions in North and South Carolina and Georgia have experienced rapid population growth over the last 10 years. It is expected that the growth will continue over the next 10 years. This has motivated many of the large grocery store chains to build new stores in the region. The Kelley's Super Grocery Stores Inc. chain is no exception. The director of planning for Kelley's Super Grocery Stores wants to study adding more stores in this region. He believes there are two main factors that indicate the amount families spend on groceries. The first is their income and the other is the number of people in the family. The director gathered the following sample information.

Family	Food	Income	Size
1	\$5.04	\$73.98	4
2	4.08	54.90	2
3	5.76	94.14	4
⋮	⋮	⋮	⋮
23	4.56	38.16	3
24	5.40	43.74	7
25	4.80	48.42	5

Food and income are reported in thousands of dollars per year, and the variable size refers to the number of people in the household.

- Develop a correlation matrix. Do you see any problems with multicollinearity?
 - Determine the regression equation. Discuss the regression equation. How much does an additional family member add to the amount spent on food?
 - What is the value of R^2 ? Can we conclude that this value is greater than 0?
 - Would you consider deleting either of the independent variables?
 - Plot the residuals in a histogram. Is there any problem with the normality assumption?
 - Plot the fitted values against the residuals. Does this plot indicate any problems with homoscedasticity?
27. **FILE** An investment advisor is studying the relationship between a common stock's price to earnings (P/E) ratio and factors that she thinks would influence it. She has the following data on the earnings per share (EPS) and the dividend percentage (Yield) for a sample of 20 stocks.

Stock	P/E	EPS	Yield
1	20.79	\$2.46	1.42
2	3.03	2.69	4.05
3	44.46	-0.28	4.16
⋮	⋮	⋮	⋮
18	30.21	1.71	3.07
19	32.88	0.35	2.21
20	15.19	5.02	3.50

- Develop a multiple linear regression with P/E as the dependent variable.
- Are either of the two independent variables an effective predictor of P/E?
- Interpret the regression coefficients.
- Do any of these stocks look particularly undervalued?
- Plot the residuals and check the normality assumption. Plot the fitted values against the residuals.
- Does there appear to be any problems with homoscedasticity?
- Develop a correlation matrix. Do any of the correlations indicate multicollinearity?

- d. Conduct an individual test to determine whether any of the independent variables can be dropped.
- e. Rerun the regression equation, using only the independent variables that are significant. How much more does a man earn per month than a woman? Does it make a difference whether the employee has a management or engineering position?
26. **FILE** Many regions in North and South Carolina and Georgia have experienced rapid population growth over the last 10 years. It is expected that the growth will continue over the next 10 years. This has motivated many of the large grocery store chains to build new stores in the region. The Kelley's Super Grocery Stores Inc. chain is no exception. The director of planning for Kelley's Super Grocery Stores wants to study adding more stores in this region. He believes there are two main factors that indicate the amount families spend on groceries. The first is their income and the other is the number of people in the family. The director gathered the following sample information.

Family	Food	Income	Size
1	\$5.04	\$73.98	4
2	4.08	54.90	2
3	5.76	94.14	4
⋮	⋮	⋮	⋮
23	4.56	38.16	3
24	5.40	43.74	7
25	4.80	48.42	5

Food and income are reported in thousands of dollars per year, and the variable size refers to the number of people in the household.

- a. Develop a correlation matrix. Do you see any problems with multicollinearity?
- b. Determine the regression equation. Discuss the regression equation. How much does an additional family member add to the amount spent on food?
- c. What is the value of R^2 ? Can we conclude that this value is greater than 0?
- d. Would you consider deleting either of the independent variables?
- e. Plot the residuals in a histogram. Is there any problem with the normality assumption?
- f. Plot the fitted values against the residuals. Does this plot indicate any problems with homoscedasticity?
27. **FILE** An investment advisor is studying the relationship between a common stock's price to earnings (P/E) ratio and factors that she thinks would influence it. She has the following data on the earnings per share (EPS) and the dividend percentage (Yield) for a sample of 20 stocks.

Stock	P/E	EPS	Yield
1	20.79	\$2.46	1.42
2	3.03	2.69	4.05
3	44.46	-0.28	4.16
⋮	⋮	⋮	⋮
18	30.21	1.71	3.07
19	32.88	0.35	2.21
20	15.19	5.02	3.50

- a. Develop a multiple linear regression with P/E as the dependent variable.
- b. Are either of the two independent variables an effective predictor of P/E?
- c. Interpret the regression coefficients.
- d. Do any of these stocks look particularly undervalued?
- e. Plot the residuals and check the normality assumption. Plot the fitted values against the residuals.
- f. Does there appear to be any problems with homoscedasticity?
- g. Develop a correlation matrix. Do any of the correlations indicate multicollinearity?

conclude that, based on vehicle counts, the US 17 store has more business volume than the SC 707 store?

Date	Time	Vehicle Count	
		US 17	SC 707
Oct 2	20:52	4	9
Oct 11	19:30	5	7
Oct 15	22:08	9	12
Oct 19	11:42	4	5
Oct 25	15:32	10	8
Oct 26	11:02	9	15
Nov 3	11:22	13	7
Nov 5	19:09	20	3
Nov 8	15:10	15	14
Nov 9	13:18	15	11
Nov 15	22:38	13	11
Nov 17	18:46	16	12
Nov 21	15:44	17	8
Nov 22	15:34	15	3
Nov 27	21:42	20	6
Nov 29	9:57	17	13
Nov 30	17:58	5	9
Dec 3	19:54	7	13
Dec 15	18:20	11	6
Dec 16	18:25	14	15
Dec 17	11:08	8	8
Dec 22	21:20	10	3
Dec 24	15:21	4	6
Dec 25	20:21	7	9
Dec 30	14:25	19	4

46. **FILE** A goal of financial literacy for children is to learn how to manage money wisely. One question is: How much money do children have to manage? A recent study by Schnur Educational Research Associates randomly sampled 15 children between 8 and 10 years old and 18 children between 11 and 14 years old and recorded their monthly allowance. Is it reasonable to conclude that the mean allowance received by children between 11 and 14 years is more than the allowance received by children between 8 and 10 years? Use the .01 significance level. What is the p -value?

8-10 Years	11-14 Years	8-10 Years	11-14 Years
26	49	26	41
33	44	25	38
30	42	27	44
26	38	29	39
34	39	34	50
26	41	32	49
27	39		41
27	38		42
30	38		30

conclude that, based on vehicle counts, the US 17 store has more business volume than the SC 707 store?

Date	Time	Vehicle Count	
		US 17	SC 707
Oct 2	20:52	4	9
Oct 11	19:30	5	7
Oct 15	22:08	9	12
Oct 19	11:42	4	5
Oct 25	15:32	10	8
Oct 26	11:02	9	15
Nov 3	11:22	13	7
Nov 5	19:09	20	3
Nov 8	15:10	15	14
Nov 9	13:18	15	11
Nov 15	22:38	13	11
Nov 17	18:46	16	12
Nov 21	15:44	17	8
Nov 22	15:34	15	3
Nov 27	21:42	20	6
Nov 29	9:57	17	13
Nov 30	17:58	5	9
Dec 3	19:54	7	13
Dec 15	18:20	11	6
Dec 16	18:25	14	15
Dec 17	11:08	8	8
Dec 22	21:20	10	3
Dec 24	15:21	4	6
Dec 25	20:21	7	9
Dec 30	14:25	19	4

46. **FILE** A goal of financial literacy for children is to learn how to manage money wisely. One question is: How much money do children have to manage? A recent study by Schnur Educational Research Associates randomly sampled 15 children between 8 and 10 years old and 18 children between 11 and 14 years old and recorded their monthly allowance. Is it reasonable to conclude that the mean allowance received by children between 11 and 14 years is more than the allowance received by children between 8 and 10 years? Use the .01 significance level. What is the p -value?

8-10 Years	11-14 Years	8-10 Years	11-14 Years
26	49	26	41
33	44	25	38
30	42	27	44
26	38	29	39
34	39	34	50
26	41	32	49
27	39		41
27	38		42
30	38		30