driver wants to determine if these calculations are different. MPGDIFF

Fill-up	1	2	3	4	5	6	7	8	9	10
Computer	41.5	50.7	36.6	37.3	34.2	45.0	48.0	43.2	47.7	42.2
Driver	36.5	44.2	37.2	35.6	30.5	40.5	40.0	41.0	42.8	39.2
Fill-up	11	12	13	14	15	16	17	18	19	20
Computer	43.2	44.6	48.4	46.4	46.8	39.2	37.3	43.5	44.3	43.3
Driver	38.8	44.5	45.4	45.3	45.7	34.2	35.2	39.8	44.9	47.5

(a) Consider the driver's mpg calculations as the explanatory variable. Plot the data and describe the relationship. Are there any outliers or unusual values? Does a linear relationship seem reasonable?

(b) Run the simple linear regression and state the least-squares regression line.

(c) Summarize the results. Does it appear that the computer and driver calculations are the same? Explain.

10.31 Gambling and alcohol use by first-year college students. Gambling and alcohol use are problematic behaviors for many college students. One study looked at 908 first-year students from a large northeastern university.[15] Each participant was asked to fill out the 10-item Alcohol Use Disorders Identification Test (AUDIT) and a 7-item inventory used in prior gambling research among college students. AUDIT assesses alcohol consumption and other alcohol-related risks and problems (a higher score means more risks). A correlation of 0.29 was reported between the frequency of gambling and the AUDIT score.

(a) What percent of the variability in AUDIT score is explained by frequency of gambling?

(b) Test the null hypothesis that the correlation between the gambling frequency and the AUDIT score is zero.

(c) The sample in this study represents 45% of the students contacted for the online study. To what extent do you think these results apply to all first-year students at this university? To what extent do you think these results apply to all first-year students? Give reasons for your answers.

10.32 Predicting water quality. The index of biotic integrity (IBI) is a measure of the water quality in streams. IBI and land use measures for a collection of streams in the Ozark Highland ecoregion of Arkansas were collected as part of a study.[16] Table 10.4 gives the data for IBI, the percent of the watershed that was forest, and the area of the watershed in square kilometers for streams in the original sample with watershed area less than or equal to 70 km^2. IBI

(a) Use numerical and graphical methods to describe the variable IBI. Do the same for area. Summarize your results.

(b) Plot the data and describe the relationship between IBI and area. Are there any outliers or unusual patterns?

(c) Give the statistical model for simple linear regression for this problem.

(d) State the null and alternative hypotheses for examining the relationship between IBI and area.

(e) Run the simple linear regression and summarize the results.

(f) Obtain the residuals and plot them versus area. Is there anything unusual in the plot?

(g) Do the residuals appear to be approximately Normal? Give reasons for your answer.

(h) Do the assumptions for the analysis of these data using the model you gave in part (c) appear to be reasonable? Explain your answer.

10.33 More on predicting water quality. The researchers who conducted the study described in the previous exercise also recorded the percent of the watershed area that was forest for each of the streams.

TABLE 10.4 Watershed Area (km^2), Percent Forest, and Index of Biotic Integrity

Area	Forest	IBI	Area	Forest	IBI	Area	Forest	IBI	Area	Forest	IBI	Area	Forest	IBI
21	0	47	29	0	61	31	0	39	32	0	59	34	0	72
34	0	76	49	3	85	52	3	89	2	7	74	70	8	89
6	9	33	28	10	46	21	10	32	59	11	80	69	14	80
47	17	78	8	17	53	8	18	43	58	21	88	54	22	84
10	25	62	57	31	55	18	32	29	19	33	29	39	33	54
49	33	78	9	39	71	5	41	55	14	43	58	9	43	71
23	47	33	31	49	59	18	49	81	16	52	71	21	52	75
32	59	64	10	63	41	26	68	82	9	75	60	54	79	84
12	79	83	21	80	82	27	86	82	23	89	86	26	90	79
16	95	67	26	95	56	26	100	85	28	100	91			

These data are also given in Table 10.4. Analyze these data using the questions in the previous exercise as a guide. IBI

10.34 Comparing the analyses. In Exercises 10.32 and 10.33, you used two different explanatory variables to predict IBI. Summarize the two analyses and compare the results. If you had to choose between the two explanatory variables for predicting IBI, which one would you prefer? Give reasons for your answer. IBI

10.35 How an outlier can affect statistical significance. Consider the data in Table 10.4 and the relationship between IBI and the percent of watershed area that was forest. The relationship between these two variables is almost significant at the 0.05 level. In this exercise you will demonstrate the potential effect of an outlier on statistical significance. Investigate what happens when you decrease the IBI to 0.0 for (1) an observation with 0% forest and (2) an observation with 100% forest. Write a short summary of what you learn from this exercise. IBI

10.36 Predicting water quality for an area of 40 km^2. Refer to Exercise 10.32. IBI

(a) Find a 95% confidence interval for the mean response corresponding to an area of 40 km^2.

(b) Find a 95% prediction interval for a future response corresponding to an area of 40 km^2.

(c) Write a short paragraph interpreting the meaning of the intervals in terms of Ozark Highland streams.

(d) Do you think that these results can be applied to other streams in Arkansas or in other states? Explain why or why not.

10.37 Compare the predictions. Consider Case 37 in Table 10.4 (8th row, 2nd column). For this case the area is 10 km^2 and the percent forest is 63%. A predicted index of biotic integrity based on area was computed in Exercise 10.32, while one based on percent forest was computed in Exercise 10.33. Compare these two estimates and explain why they differ. Use the idea of a prediction interval to interpret these results. IBI

10.38 Reading test scores and IQ. In Exercise 2.33 (page 100) you examined the relationship between reading test scores and IQ scores for a sample of 60 fifth-grade children. READIQ

(a) Run the regression and summarize the results of the significance tests.

(b) Rerun the analysis with the four possible outliers removed. Summarize your findings, paying particular attention to the effects of removing the outliers.

10.39 Leaning Tower of Pisa. The Leaning Tower of Pisa is an architectural wonder. Engineers concerned about the tower's stability have done extensive studies of its increasing tilt. Measurements of the lean of the tower over time provide much useful information. The following table gives measurements for the years 1975 to 1987. The variable "lean" represents the difference between where a point on the tower would be if the tower were straight and where it actually is. The data are coded as tenths of a millimeter in excess of 2.9 meters, so that the 1975 lean, which was 2.9642 meters, appears in the table as 642. Only the last two digits of the year were entered into the computer.[17] PISA

Year	75	76	77	78	79	80	81	82	83	84	85	86	87
Lean	642	644	656	667	673	688	696	698	713	717	725	742	757

(a) Plot the data. Does the trend in lean over time appear to be linear?

(b) What is the equation of the least-squares line? What percent of the variation in lean is explained by this line?

(c) Give a 99% confidence interval for the average rate of change (tenths of a millimeter per year) of the lean.

10.40 More on the Leaning Tower of Pisa. Refer to the previous exercise. PISA

(a) In 1918 the lean was 2.9071 meters. (The coded value is 71.) Using the least-squares equation for the years 1975 to 1987, calculate a predicted value for the lean in 1918. (Note that you must use the coded value 18 for year.)

(b) Although the least-squares line gives an excellent fit to the data for 1975 to 1987, this pattern did not extend back to 1918. Write a short statement explaining why this conclusion follows from the information available. Use numerical and graphical summaries to support your explanation.

10.41 Predicting the lean in 2013. Refer to the previous two exercises. PISA

(a) How would you code the explanatory variable for the year 2013?

(b) The engineers working on the Leaning Tower of Pisa were most interested in how much the tower would lean if no corrective action was taken. Use the least-squares equation to predict the tower's lean in the year 2013. (*Note:* The tower was renovated in 2001 to make sure it does not fall down.)

(c) To give a margin of error for the lean in 2013, would you use a confidence interval for a mean response or a prediction interval? Explain your choice.

10.42 Correlation between binge drinking and the average price of beer. A recent study looked at 118 colleges to investigate the association between the binge-drinking rate and the average price for a bottle of beer at establishments within a two-mile radius of campus.[18] A correlation of -0.36 was found. Explain this correlation.

10.43 Is this relationship significant? Refer to the previous exercise. Test the null hypothesis that the correlation between the binge-drinking rate and the average price for a bottle of beer within a two-mile radius of campus is zero.