Lesson 20: Analyzing Data Collected on Two Variables

Classwork

Lessons 12–19 included several data sets that were used to learn about how two numerical variables might be related. Recall the data on elevation above sea level and the number of clear days per year for fourteen cities in the United States. Could a city's elevation above sea level by used to predict the number of clear days per year a city experiences? After observing a scatter plot of the data, a linear model (or the least-squares linear model obtained from a calculator or computer software) provided a reasonable description of the relationship between these two variables. This linear model was evaluated by considering how close the data points were to the corresponding graph of the line. The equation of the linear model was used to answer the statistical question about elevation and the number of clear days.

Several data sets were also provided to illustrate other possible models, specifically quadratic and exponential models. Finding a model that describes the relationship between two variables allows you to answer statistical questions about how the two numerical variables vary. For example, the following statistical question was posed in Lesson 13 regarding latitude and the mean number of flycatcher chicks in a nest: What latitude is best for hatching flycatcher chicks? A quadratic model of the latitude and the mean number of flycatcher chicks in a nest was used to answer this statistical question.

In Lessons 12–19, you worked with several data sets and models used to answer statistical questions. Select one of the data sets presented in Lessons 12–19, and develop a poster that summarizes how a statistical question is answered that involves two numerical variables. Your poster should include the following: a brief summary of the data, the statistical question asking the relationship between two numerical variables, a scatter plot of the data, and a brief summary to indicate how well the data fit a specific model.

After you identify one of the problems to summarize with a poster, consider the following questions to plan your poster:

1. What two variables were involved in this problem?

2. What was the statistical question? Remember, a statistical question involves data. The question also anticipates that the data will vary.

3. What model was used to describe the relationship between the two variables?

4. Based on the scatter plot of the data, was the model a good one?

5. Was the residual plot used to evaluate the model? What did the residual plot indicate about the model?

6. How would you use the model to predict values not included in the data set?

7. Does the model answer the statistical question posed?

Examples of posters involving two numerical variables can be found at the website of the American Statistical Association (www.amstat.org/education/posterprojects/index.cfm).