



Lesson 20: Analyzing Data Collected on Two Variables

Student Outcomes

- Students use data to develop a poster that involves the focus standards.
- Students construct a scatter plot of the data.
- Students analyze their data, examine the residual plot, and interpret the correlation coefficient.

Lesson Notes

MP.3

This lesson provides an opportunity to integrate the methods presented in Lessons 12–19. In this lesson, students develop a poster to explore the relationship between two numerical variables. This lesson directs students to select one of the data sets introduced in Lessons 12–19 and to summarize how well the data fit a specific model. Encourage students who may have struggled with this process to select one of the linear relationships, because understanding linearity provides the basis for understanding other models.

Instructors must decide the following prior to the lesson:

- What is the format of the poster or presentation?
 - *Creating posters may require supplies, such as poster board and markers. Large post-it paper could also work for the posters. Teachers may prefer to have students create a short presentation using PowerPoint or similar presentation software. Or, teachers may want students to develop the posters in small groups or individually.*
- How will the presentation be evaluated?
 - *A well-defined rubric is recommended; an example is available at the following location: www.amstat.org/education/posterprojects/index.cfm.*
- As rubric designs are highly dependent on the process used to complete this type of presentation, the final rubric design is the teacher's decision. To assess the poster or presentation, teachers should look at how well students addressed the following: the statistical question of how the variables presented are related, the scatter plot and the fit of a possible function to the data, an explanation of the selected model, and an answer to the statistical question.

Classwork

Read through the introductory paragraph with students. The poster or presentation will deal with numerical variables.

Lessons 12–19 included several data sets 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 be 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).

Ask the students:

- What is the difference between categorical and numerical variables? (This question reminds students that they are not selecting a data set that was summarized by a two-way frequency table.)

Recall with students some of the examples of models from Lessons 12–19 that would be excellent problems to summarize in a poster. The opening paragraphs in the student lesson refer to elevation of a city and the mean number of clear days, as well as latitude and the number of chicks in a nest from Lesson 12. Additional examples:

- Shoe length and height (Lesson 14 Example)
- Animal gestation time and longevity (Lesson 15 Example)
- Time and cost of surgery (Lesson 15 Problem Set)
- Curb weight and fuel efficiency (Lesson 16 Example)
- Athlete track times (Lesson 16 Problem Set)
- Temperature and volume (Lesson 18 Exercise)

Organize students in small groups to select one of the models from Lessons 12–19 for their presentation. Explain the chosen format for the presentation (poster, PowerPoint). Also explain whether students should develop their posters in small groups or if they are expected to develop the posters individually. The poster or presentation should include a statistical question, a scatter plot, and an evaluation that indicates how well the data fits the model. Use the seven questions in the lesson as a guide for helping students organize their presentations.

As students are working, walk around and ask the following:

- (1) What two variables are you studying? Why do you think the model connects the two variables?
 - *Students should define their variables based on the data set they selected.*
- (2) What is the statistical question you would like to answer?
 - *Again, have students reflect on the data. Ask them to describe the context that is connected to the data, possibly indicating how it might have been collected. Also ask them to indicate why the data may vary.*
- (3) How does the model describe the relationship between the two variables?
 - *Encourage students to talk about the trend: linear, quadratic, or exponential. If linear, what is the meaning of the slope and y-intercept within the context of the problem?*
- (4) Is the model appropriate based on a scatter plot of the data?
 - *Encourage students to talk about the trend; linear, quadratic, exponential. Students should also discuss estimates of residuals and how close the points are to the graph of the least-squares line.*
- (5) How are you evaluating the model selected for the data?
 - *Students can use residual plots or correlation coefficients to justify the strength and appropriateness of the model.*
- (6) What predictions could you make from the model of the relationship between the two variables?
 - *Examples from the lessons include:*
What is the estimated age of a lobster with an exterior shell length of 75 mm or less?
What elevations above sea level are predicted to have at least 100 clear days per year?

If time permits, organize a display of the posters. Encourage students to explain their posters to other students or to the entire class.

MP.3