Investigation 5

How Many Calories?

Scatterplots

Overview

This investigation begins the exploration of relationships within bivariate data by investigating errors made by students between guesses and actual values, setting the stage for the concept of a residual.

The concept of a residual and a residual plot will be used in Investigation 7 as a tool for exploring variability about the least squares regression line.

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Pre-K-12 Report*. The four components are formulate a statistical question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level B activity.

This activity is based on lessons from *Exploring Linear Relations* (Lesson 7), published by the American Statistical Association (original copyright by Dale Seymour Publications 1999). *Exploring Linear Relations* is a module in the ASA Data-Driven Mathematics Project. It is available as a free download at *www.amstat.org/asa/files/pdfs/ddmseries/ExploringLinearRelations--TeachersEdition.pdf*.

Instructional Plan

Brief Overview

» Read and discuss the scenario about obesity and caloric information.

» Formulate the statistical question: What is the typical error made by students in estimating the number of calories in bite-sized candies?

» Have students estimate the number of calories in each item and find their errors.

» Use the squares of the errors to determine who is the "best" guesser.

Hand out Student Worksheet 5.1 Guess the Calories. Have your students read the Scenario.

Scenario

The following excerpt comes from *Attacking the Obesity Epidemic: The Potential Health Benefits of Providing Nutrition Information in Restaurants* by Scot Burton, Elizabeth H. Creyer, Jeremy Kees, and Kyle Huggins. The entire article can be found at *www.ncbi.nlm.nih.gov/pmc/articles/PMC1551968*.

> Sixty-four percent of American adults are either overweight or obese, and the obesity epidemic shows few signs of weakening. Although the precise number of deaths attributable to obesity is difficult to estimate, obesity is clearly a major cause of preventable death. Not surprisingly, improving the healthfulness of the American diet has become a national health priority. The increasing prevalence of obesity-related diseases has been blamed, in part, on the increased consumption of foods prepared outside the home. Restaurant expenditures have increased consistently in

Learning Goals

» Represent data on two quantitative variables on a scatterplot and describe how the variables are related

» Develop understanding of an error

Mathematical Practices Through a Statistical Lens

MP3. Construct viable arguments and critique the reasoning of others.

Statistically proficient students use appropriate data and statistical methods to draw conclusions about a statistical question. They follow the logical progression of the statistical problem-solving process to investigate answers to a statistical question and provide insights into the research topic. They reason inductively about data, making inferences that consider the context from which the data arose. They justify their conclusions, communicate them to others (orally and in writing), and critique the conclusions of others.

Materials

Student worksheets are available at *www.statisticsteacher.org/statistics-teacher-publications/focus.*

» Student Worksheet 5.1 Guess the Calories

» Graph paper (.5 cm or .25 inch)

» "Fun" size Milky Way candy bar

Estimated Time

One 50-minute class period

Pre-Knowledge

Students should already be able to:

» Construct scatterplots

» Draw the y=x line on a scatterplot

recent decades; consumers now spend more than $400 billion annually.

Results: Survey results showed that levels of calories, fat, and saturated fat in less-healthful restaurant items were significantly underestimated by consumers. Actual fat and saturated fat levels were twice consumers' estimates and calories approached two times more than what consumers expected. In the subsequent experiment, for items for which levels of calories, fat, and saturated fat substantially exceeded consumers' expectations, the provision of nutrition information had a significant influence on product attitude, purchase intention, and choice.

Conclusions: Most consumers are unaware of the high levels of calories, fat, saturated fat, and sodium found in many menu items. Provision of nutrition information on restaurant menus could potentially have a positive impact on public health by reducing the consumption of less-healthful foods.

Formulate a Statistical Question

After students have read the scenario ask:

How well do you think you might estimate the calories in restaurant items? Have you noticed restaurants providing this information more readily? For example, some restaurants list this on their menus, such as Panera and Noodles and Co.

Explain that they will be asked to estimate the number of calories in "fun" size candy bars. Hold up a fun size Milky Way candy bar and ask students to guess the number of calories. Have them write down their guesses, and then ask a few students to share. Then share that the fun size Milky Way bar contains 80 calories.

Ask your students how close their estimates were for the number of calories.

Ask students to create a possible statistical question for this activity.

Example: *What is the typical error made by students in estimating the number of calories in bite-sized candies?*

Note: If it appears students may interpret this investigation as the worst guessers will become obese, it may be worth making the point that the study was looking for a positive relationship between knowing the calories of a food prior to consumption and making healthier choices. This does not indicate not knowing the calories in food means making unhealthy choices.

Table 5.1 Candy

Candy Item – Fun Size	Actual
Snickers	80
Skittles	80
Butterfinger	100
Kit Kat	70
M&M's Plain	73
M&M's Peanut	90
Reese's Peanut Butter Cup	110
Starburst	40
Whoppers	100
Twizzlers	50
Jolly Ranchers (3 Pieces)	70

Collect Appropriate Data

Ask students to complete problem 1 on Student Worksheet 5.1 Guess the Calories.

1. Fill in the "Guess?" column with your guesses for the number of calories in each fun size candy item.

When students have completed their Guess? column, reveal the actual number of calories (Table 5.1). Have the students complete Problem 2.

2. Fill in the "Actual" column with the actual number of calories in each fun size candy item.

Analyze the Data

Ask your students to complete Question 3.

3. How might you decide who is the best guesser in the class? Justify your answer.

After giving your students time to individually write about their methods of determining the best guesser, have them share with a partner or in small groups. Then, as a whole group, ask students to share their ideas about how they might determine who is the best guesser. Does a person need to be exactly

correct? Or just close? Is underestimating different from overestimating?

Possible answers: *Students might suggest a variety of ways to calculate who is the best guesser. Encourage discussion and collect several ideas. For example, the greatest number of guesses that match actual calories or the greatest number of guesses that were within 10 calories of the actual calories. If it doesn't come up, guide discussion around whether it matters how far off someone was.*

Explain that we are going to create a visual display of their data to help determine who is the best guesser. Distribute graph paper to each student. Give students time to answer questions 4 and 5.

4. Create a scatterplot of your data on graph paper, plotting the actual calories on the x-axis and your guess on the y-axis.

Sample answer: *Figure 5.1*

5. Describe the relationship between your guesses and the actual calories in each candy item.

Possible answer: *As the actual calories increase, my guessed calories increased. I tended to underestimate the actual calories. For the Butterfinger, my guess for number of calories was much higher than the actual number of calories.*

Table 5.2 Sample Data

Candy Item – Fun Size	Actual	Guess?
Snickers	80	90
Skittles	80	75
Butterfinger	100	150
Kit Kat	70	50
M&M's Plain	73	50
M&M's Peanut	90	85
Reese's Peanut Butter Cup	110	90
Starburst	40	25
Whoppers	100	50
Twizzlers	50	30
Jolly Ranchers (3 Pieces)	70	50

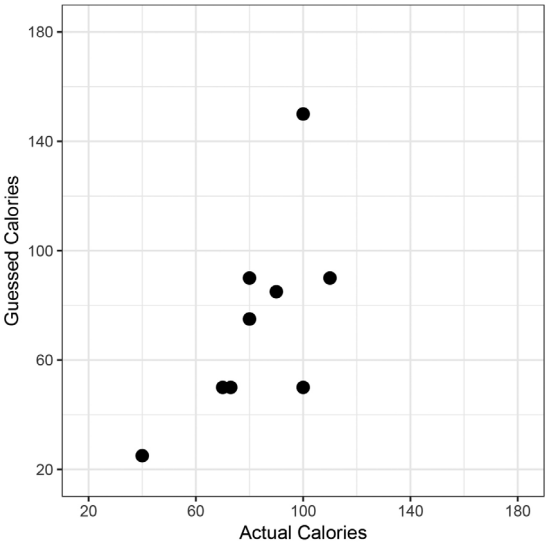

Figure 5.1: Scatterplot of guessed and actual calories

Use one of the student's scatterplots as an example or show students a scatterplot with the x-axis labeled Actual Calories and the y-axis labeled Guessed Calories. Ask what a point on this graph represents. For example, the point (100, 75) represents a candy item that has 100 calories and was guessed to have 75 calories. Refer to the study from the beginning of the lesson—if someone often underestimated the number of calories in the candy items, how would that appear in a scatterplot of the data?

The points would be closer to the x-axis since the guessed calories would be lower than the actual calories.

Ask your students to answer Question 6,

6. What would the scatterplot look like if someone had guessed the correct actual calories in each candy item?

Possible answer: *The points would make a straight line. If it doesn't come up, ask for the equation of the line, which is y = x, the line*

where all the y-values (guessed calories) are the same as the x-values (actual calories).

Have students draw the y = x line on their scatterplots. Ask what it means for a point to be on the line, below the line, and above the line.

Possible answers: *If a point is on the line, the guess matches the actual calories. If a point is below the line, the guess is lower than the actual calories. If a point is above the line, the guess is higher than the actual calories.*

Ask students to answer Question 7, and then have some students share their scatterplots and explain the type of guesser their graph shows.

7. Describe the type of guesser your scatterplot shows. Explain.

Possible answer: *Based on my scatterplot (Figure 5.2), I tended to underestimate the calories in the fun size items as shown by most of my dots being below the line y = x.*

Ask students how this line might help us determine who is a better guesser.

Possible answer: *Students might suggest a person whose points are closer to the line is a better guesser.*

How might we measure the distance from the line?

Possible answer: *Students might suggest measuring the perpendicular distances, horizontal distances, vertical distances, or other methods.*

Explain that we are going to use the vertical distances, as this is a convention used in statistics. What do the vertical distances represent?

Possible answer: *The difference between the guessed calories (y-value of the point) and the actual calories (y-value on the line).*

What could we call these differences?

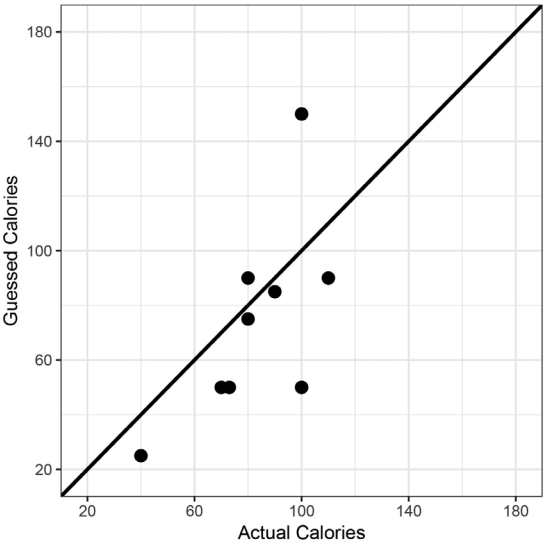

Figure 5.2: Scatterplot of someone who underestimated the calories in the candy.

Possible answer: *These differences between the guessed calories and actual calories for each candy item are the amounts of error for each candy item.*

Note: You should not use the term "residual." This activity is setting the stage for understanding the concept of the residual in Investigation 7.

The values for the vertical distances represent the error for each guess. Using a student's scatterplot, choose a point above the line and draw the vertical distance between the point and the line y = x. The example here shows the point (100, 150), which means the difference between the guessed calories (150) minus the actual calories (100) is 50, thus the error is 50 calories.

If an error is 10, how could we determine if the person overestimated or underestimated? If no students make a suggestion, explain that positives and negatives are used to determine this. The error is positive if the point is above the line, and the error is negative if the point is below the line. In this situation, points above

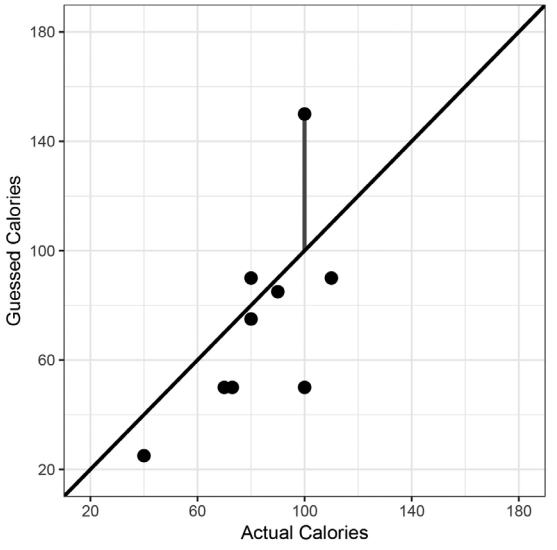

Figure 5.3: Scatterplot showing the difference between the guessed calories and actual calories for each candy

Table 5.3 Errors Based on Sample Data

Candy Item – Fun Size	Actual	Guess?	Errors
Snickers	80	90	10
Skittles	80	75	-5
Butterfinger	100	150	50
Kit Kat	70	50	-20
M&M's Plain	73	50	-23
M&M's Peanut	90	85	-5
Reese's Peanut Butter Cup	110	90	-20
Starburst	40	25	-15
Whoppers	100	50	-50
Twizzlers	50	30	-20
Jolly Ranchers (3 Pieces)	70	50	-20

the line are overestimates and positive errors; points below the line are underestimates and negative errors. It might be helpful to show another example on a student's graph.

Ask what an error of 0 represents.

Answer: *The guess and the actual value were the same.*

Have your students answer questions 8 and 9.

8. On your scatterplot, draw the y=x line. Then draw the vertical distances representing the "errors" on your scatterplot.
9. On the table of guesses and actual number of calories, add a third column labeled "Errors" and calculate the errors (guess minus actual) for each candy item. Find the sum of the errors.

Answer based on the given example (Table 5.3).

Ask students how we might use the errors to help us determine who is the best guesser. Encourage discussion. Students might suggest how many are close to 0 or +/- 10 calories or adding up the errors. Propose that the best guesser is one who has a sum of errors close to zero. After students have found their sum, ask who the best guesser was using this method. Ask students if they have any concerns about using the sum of the errors.

Possible answer: *This would not be a good method since someone could have a very high error (extreme overestimate) balanced by a very low error (extreme underestimate).*

Ask students how we might handle values that are negative when we might like them to be positive.

Possible answer: *Students will most likely say take the absolute value.*

Have your students answer Question 10.

10. On the table of guesses and actual number of calories, add a fourth column labeled "Absolute Value" and complete the column. Find the sum of the absolute values.

Note: Finding the absolute values and the sum of the absolute values could be connected to earlier learning of the Mean Absolute Deviation (MAD) from middle school.

Now ask again who is the best guesser. The best guesser would have the lowest sum of the absolute value of errors.

Ask students for another way we might handle values that are negative when we might like them to be positive.

Possible answer: *We could square the values.*

Ask students where else squaring has been used in statistics to "handle" negatives.

Answer: *Squaring was used when calculating deviations from the mean for standard deviation.*

Ask how this would help determine who the best guesser is.

Possible answer: *The best guesser would have the lowest sum of the squares of the errors, though this will usually be much higher than the sum of the absolute value of the errors.*

Note: If all the errors are less than 1, then the sum of squares will be less.

Have your students answer Question 11.

11. On the table of guesses and actual number of calories, add a fifth column labeled "Squares" and complete the column. Find the sum of the squares.

Ask again who the best guesser is. Did this answer change from the best guesser based on absolute values of the errors?

Interpret the Results in the Context of the Original Question

In groups of four, ask your students to complete problems 12 to 15. Then discuss.

12. Compare your results from Question 11 with the other students in your group. Who in your group was the best guesser of calories? Justify your answer.

Possible answer: *I know ____ is the best guesser because his/her sum of the squared errors is the lowest. This means his/her guesses were closest to the line y=x, which would be the line created if all the guesses were correct.*

13. Using the scatterplots and analysis of the errors, answer the statistical question: What is the typical error made by students in estimating the number of calories in bite-sized candies?

Answers will vary depending on the class results. Reference whether the students tended to overestimate or underestimate. Also refer to the usual number of calories their estimate was off.

14. How do these results relate to the study results?

Possible answer: *Based on the results from the class, either support or do not support the claim in the study that people tend to underestimate the number of calories in restaurant items.*

15. Why might finding errors be important when looking at data?

Possible answer: *Errors can help determine how close guesses are to the actual data values and who is the best guesser from many guessers.*

Additional Ideas

Do the same investigation steps, but instead of using the candy items, guess the calories (or fat grams) of fast food items from a particular restaurant or several restaurants. Students can find this information online.

Do the same investigation steps, but instead of using the candy items, guess the ages of various celebrities such as actors, sports figures, or politicians (international, national, and/or local) students would know. It can be helpful to have recent pictures of the celebrities to show. Make sure to have variety in ages.

Exit Ticket

1. The Price Is Right - Cliffhanger Game: "The contestant bids on three small prizes. For every dollar the contestant is away from the actual prices, a mountain climber takes one step up a mountain. If the mountain climber does not exceed 25 steps after the contestant has bid on all three prizes, then the contestant wins a bonus prize."
   Source: www.priceisright.com/games

Here are some items and the prices the contestants bid. Who won the bonus prize? Who was the best price guesser? Who was the worst guesser? Justify your answer.

Jenna		
Item	**Bid**	**Actual**
Passport Holder	12	16
Toaster	35	22
Coffee Pot	35	40
Lindsey		
Item	**Bid**	**Actual**
Inflatable Pool Lounge Chair	15	20
Electronic Piggy Bank	37	30
Pet Brush and Accessories	27	40
Debra		
Item	**Bid**	**Actual**
Liquid Measuring Cup	5	15
Electric Egg Cooker	7	22
Whipped Cream Dispenser	6	26
Kristen		
Item	**Bid**	**Actual**
Steam Iron	25	22
Electric Heater Fan	35	35
Milkshake Drink Mixer	50	49

Exit Ticket Cont.

Exit Ticket Answer:

Jenna					
Item	**Bid**	**Actual**	**Error**	**Absolute Value of Error**	**Squared Error**
Passport Holder	12	16	-4	4	16
Toaster	35	22	13	13	169
Coffee Pot	35	40	-5	5	25
Sum	22	210			
Lindsey					
Item	**Bid**	**Actual**	**Error**	**Absolute Value of Error**	**Squared Error**
Inflatable Pool Lounge Chair	15	20	-5	5	25
Electronic Piggy Bank	37	30	7	7	49
Pet Brush and Accessories	27	40	-13	13	169
Sum	25	243			
Debra					
Item	**Bid**	**Actual**	**Error**	**Absolute Value of Error**	**Squared Error**
Liquid Measuring Cup	5	15	-10	10	100
Electric Egg Cooker	7	22	-15	15	225
Whipped Cream Dispenser	6	26	-20	20	400
Sum	45	725			
Kristen					
Item	**Bid**	**Actual**	**Error**	**Absolute Value of Error**	**Squared Error**
Steam Iron	25	22	3	3	9
Electric Heater Fan	35	35	0	0	0
Milkshake Drink Mixer	50	49	1	1	1
Sum	4	10			

Jenna, Lindsey, and Kristen all won the bonus prize because their sum of the absolute values of the errors did not exceed 25. The best guesser is Kristen, since her sum of squared errors is only 10. The worst guesser is Debra, as her sum of squared errors is 725. Debra underestimated all the prices by quite a bit.

Further Explorations and Extensions

1. The Price Is Right - Bargain Game: "Two prizes are shown to the contestant. Each prize displays a bargain price that is below its actual retail price. If the contestant selects the bigger bargain (the prize with the bargain price that is farther from the actual retail price), then the contestant wins both prizes."
   Source: www.priceisright.com/games

Sketch a scatterplot that would depict what several prizes and their bargain prices might look like, as well as the line y = x.

Example video: *www.youtube.com/watch?v=crzYmKNvISg*

Example from video: Which price is the bigger bargain? Billiards table at $1600 or 55" HDTV at $2649?

Billiards table: bargain price: $1600; actual price: $2600

55" HDTV: bargain price: $2649; actual price: $3149

Answers: *Figure 5.4 and Figure 5.5*

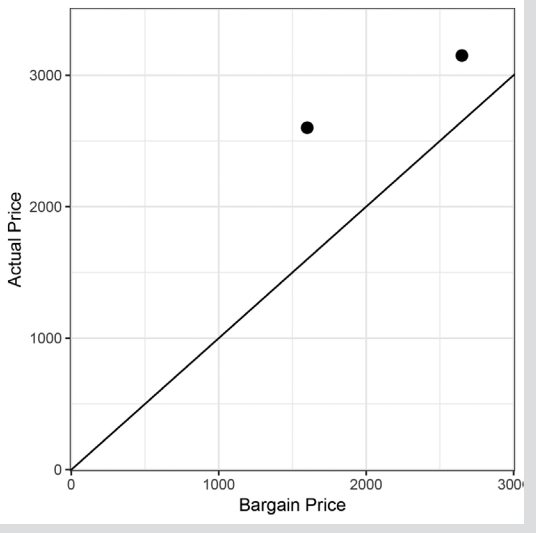

Figure 5.4

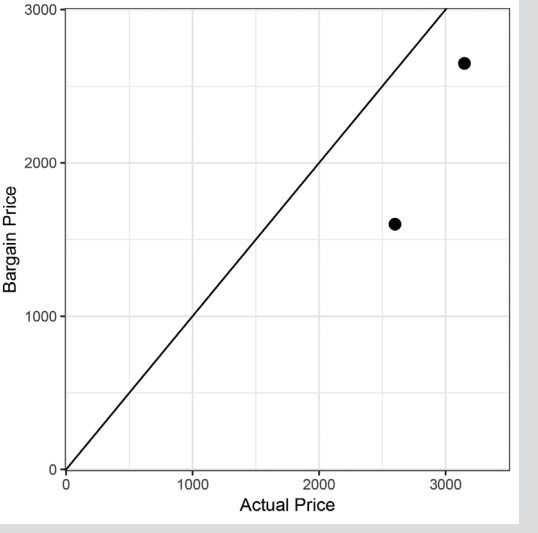

Figure 5.5