INTRODUCTION

ACTIVITIES

PROJECTS

LINE 'EM UP

Prerequisites

Observed y: y
Observed x: x
Predicted y: $\hat{y}$
Residual: $y - \hat{y}$

Overview

In teaching students about the least squares regression line, it is helpful for students to try several lines and see how the value of $\sum(y-\hat{y})^2$ changes from line to line. After discovering the concept of a line of best fit and $y-\hat{y}$, it is important for students to investigate several lines, evaluate the fit of the line to the data and then see how the least-squares regression line meets the "best fit" criteria.

Directions

Pick four points in a somewhat linear relationship (but not all on the same line). Have students plot these points on graph paper using the whole sheet (or at least ½ a sheet) so the units along the axes are large and the points are clearly distinct.

Pick any two points and have students use one of the colored pencils to draw the line connecting the points on the graph paper. Determine the equation of the line. Students should use this equation to determine the predicted y value for the x value of each point. Results can be recorded on the accompanying worksheets.

The calculations could be done using the list features of the TI-83 calculator, but it is important to complete the first set of calculations manually so that students get a clear understanding of how each value in the worksheet chart is determined.

Materials

1 sheet of graph paper per student
Several colored pencils or markers
Ruler/straight edge
TI-83 calculator

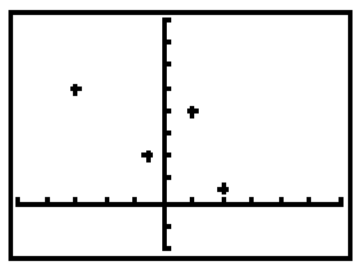
Repeat these steps for at least three possible lines. Some lines will obviously not be good choices, but that can help demonstrate how large the value of $\sum(y-\hat{y})^2$ is when the line is not a good choice.

Finally, the equations for the least-squares regression line should be determined and the tables on the worksheet completed. Use any method you wish to determine the equation (regression function on a calculator, or the formulas for calculating values for a and b).

Analyzing the Data activities help students draw conclusions about the appropriateness of each line and how it meets the "best fit" criteria.

Below is an example of four possible points and three equations.

(-3, 5) (-0.5, 2) (1, 4) (2, 0.5)



Two Selected Points	Equation	$\sum(y-\hat{y})^2$
(-3, 5) (-0.5, 2)	$y = \left(-\frac{6}{5}\right)(x+3)+5$	14.69
(-3, 5) (1, 4)	$y = \left(-\frac{1}{4}\right)(x+3)+5$	16.203
(-3, 5) (2, 0.5)	$y = \left(-\frac{9}{10}\right)(x+3)+5$	7.3225
Least-Squares Line	$y = 2.79075 + -0.67401x$	5.74229

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ___________________________

Collecting the Data

1. Write down the four points used in this activity.

   (,) (,) (,) (,)

2. Plot the four points on the graph paper. Use the entire piece of graph paper.

3. Pick two points. (,) (,)
   On the graph paper, use a colored pencil to draw the line joining the two selected points.

   Write the equation of the line joining the two points. ________________

   Fill in this table.

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$

   Find the sum of the column $(y-\hat{y})^2$. ______________________

4. Pick two new points. (,) (,)
   On the graph paper, use another colored pencil to draw the line joining the two selected points.

   Write the equation of the line joining the two points. ________________

   Fill in this table.

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$

   Find the sum of the column $(y-\hat{y})^2$. ______________________

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ____________________

5. Pick two new points. (,) (,)
   On the graph paper, use another colored pencil to draw the line joining the two selected points.

   Write the equation of the line joining the two points. ________________

   Fill in this table.

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$

   Find the sum of the column $(y-\hat{y})^2$. ______________________

6. Pick two new points. (,) (,)
   On the graph paper, use another colored pencil to draw the line joining the two selected points.

   Write the equation of the line joining the two points. ________________

   Fill in this table.

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$

   Find the sum of the column $(y-\hat{y})^2$. ______________________

7. Following your teacher's directions, find the equation of the least-

   squares regression line. ____

   Draw the least-squares regression line and clearly label it on the graph paper.

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ____________________

Fill in this table.

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$

Find the sum of the column $(y-\hat{y})^2$. ____________________

Analyzing the Data

8. Fill in the chart below using the information from questions 3-7.

Equation	Equation	$\sum(y-\hat{y})^2$
Equation 1		
Equation 2		
Equation 3		
Equation 4		
Least-Squares Regression		

9. Looking at the graphs, which line appears to best fit the trends in the data? Why? (Use only graphical reasons.)

10. Looking at the values of $\sum(y-\hat{y})^2$, which equation had the lowest sum? Is it the equation you selected in the previous question?

11. Explain the two features of the line of best fit that you have discovered in this activity.

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

IT'S IN THE BAG

Overview
Many times in dealing with sample data, we cannot always compute precisely a characteristic of the population such as *p*, the population proportion. Instead we may give an interval of values that we feel comfortable with in predicting a likely value for the parameter. Such an interval is called a confidence interval. A *confidence interval* for a population characteristic is the interval of plausible values for the characteristic. It is constructed so that, with a specified degree of confidence, the value of the characteristic will be captured inside the interval.

A related concept is the *confidence level*. It tells us the degree of confidence in the method used to construct the estimate of the interval. If we construct a 95% confidence interval, it means that our sampling method would capture the true population characteristic about 95% of the time.

This activity is designed to help students see what a confidence interval is. It will also show them that the process produces a "fuzzy" set of data so that more than one confidence interval can capture the true mean.

Materials
1 small paper sack for each student (If you have a very small class, you may want to use 2 sacks for each student.)
A large supply of colored centimeter cubes, colored disks, M&M's, or Skittles. You could also use uncooked red and white kidney beans.

Directions
Prior to class, prepare a sack for each student. Randomly label each sack with a letter or number. Decide on a specific color, A, to be sampled. Place a different number of color A objects in each sack, making sure that each sack contains a total of 25 objects. Place 13 color A objects in each of three sacks. Record which sacks contained the 13 color A objects (52% of the total).

As the discussion begins, hand each student a sack and tell them not to look inside. Tell them that all the sacks contain objects, but that you are only interested in color A objects. Ask them to guess who has the sack that contains 52% of the color A objects. (Don't tell them there are three such sacks yet.) Guide the resulting discussion to develop the concept that it is impossible to know without information about what is in each sack. The process we will use to gain this information is random sampling.

Hand out the worksheet and have the students follow the listed directions. Remind them that sampling with replacement means to return the cube to the sack and shake the sack before drawing another cube.

While the students are gathering data, in two different locations on the board, draw a horizontal axis and label it from 0 to 1 with increments of 0.1. On each axis, draw a vertical line at $x = 0.52$ to represent the desired proportion. After students have established their intervals, have them come up to the board and on the first axis, draw a horizontal line segment to represent their confidence interval, labeling the line with the letter on the sack.

Ask the students which sacks they think are unlikely candidates to have the correct proportion and have them explain their reasoning. Next, ask which sacks are likely candidates. Ask how sure they are of their choices. Let them defend their decisions. Have them complete samples 5 and 6. Let them draw a new confidence interval on the other axis based on how confident they are as to the correct proportion in their sack. These new intervals are likely to differ in size from the first interval they drew. Usually, the students draw shorter intervals because the additional data gives them more information about the objects in their sack.

From this introductory activity, you can explain what a confidence interval is and how to compute with a specified degree of confidence.

Follow-up Activity
Have each student in the class poll 15 students at random during the day about a question of interest to your students. Let each student compute the proportion of interest and confidence interval. Then have them graph the intervals on the board. As a class, decide what the most likely value for the desired proportion is.

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ____________________________

I. Collecting the Data

In the sack you received are 25 objects. You are to determine what proportion of the objects are the color your teacher listed. You will do this by sampling with replacement. Each time you draw an object from the sack, record whether it is the correct color or not. Then place the object back in the sack. To get a large enough sample, you are to draw an object 10 times from the sack. You will repeat the process for a total of 4 samples.

Sample 1

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

Sample 2

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

Sample 3

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

Sample 4

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ____________________________

II. Analyzing the Data

1. What was the smallest proportion of the correct color that you had? ______

2. What was the greatest proportion of the correct color that you had? ______

3. Based on the data you collected, what would you guess is a likely interval for the correct proportion in your sack? ______ to _______ Explain your answer.

4. Do you think it is likely that your sack was the one whose proportion was 0.52? How confident are you in your decision? Explain.

5. Compare your results with others in the class. Which sacks do you think could qualify as the one with a proportion of 0.52?

6. Is it possible that none of the sacks you just listed had the correct proportion of 0.52? Why?

7. Compare your results with others in the class. Which sacks do you think don't qualify as the one with a proportion of 0.52?

8. Is it possible that one of the sacks you just listed had the correct proportion of 0.52?

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.

Name: ____________________

9. How likely do you think it is? Why?

10. What could you do statistically to narrow your choices of the correct sack?

11. Repeat the sampling process two more times.

Sample 5

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

Sample 6

Drawing #	1	2	3	4	5	6	7	8	9	10	Proportion Correct Color
Correct Color? Y/N											

12. How likely do you think it is now that your sack contains the correct proportion?

13. Compare your new results with the rest of the class.
Which sacks do you now think are likely to have the correct proportion?

14. Which one sack do you think is **most** likely to contain the correct proportion?

15. How can you be totally sure?

Copyright © 2000 William K. Bradford Publishing Company. All rights reserved. Permission to reproduce this master is granted to registered purchasers for their classroom use. Printed in the United States of America.