

Lesson 3: Residuals and Coefficient of Determination

Name _____ Period 1 2 3 4 5 6 7 8

In this activity you will take a data set and input into Core Math Tools to learn how to use the program and review from Lesson 1. We will also focus on finding, interpreting, and graphing best fit lines, and residuals, and learning about the coefficient of determination.

The following data set is taken from The Practice of Statistics Fourth Edition, Starnes Yates and Moore.


Over the last few years, many people have gone on “low-carb” diets while others have tried “low-fat” diets. Here are the carbohydrate contents in grams (g) and fat contents in grams for nine different types of hamburgers at McDonald’s:

Type	Carbs(g)	Fat(g)
Hamburger	31	9
Cheeseburger	33	12
Double Cheeseburger	34	23
Quarter Pounder	37	19
Quarter Pounder with cheese	40	26
Dbl Qtr Pound w/cheese	40	42
Big Mac	45	29
Big N'Tasty	37	24
Big N'Tasty with cheese	38	28

Directions:

1. Go to Core Math Tools, click on Data Analysis and input your data from the above table.
2. Put the Carbs in column A and Fat in column B.
3. Go to Edit and click on column name and label your x & y axis accordingly. You will need to do this for each column. To get the column make sure a cell in the column is highlighted.




4. Now click on the scatter plot button . Make sure your horizontal axis has carbs highlighted and your vertical axis has fat highlighted. Click OK and your scatter plot should appear.
5. Describe the relationship between the two variables in form, direction, and strength.

6. Estimate what r is by going to “Options” and clicking on Guess Correlation_____.

Now click on “calculate r ”. _____ Were you correct?_____

7. What does the r value measure between the variables?

8. Click on the  (green line with pencil) button. A line will appear with its equation in the upper right-hand corner of the scatter plot. There will be 3 boxes on your line. Clicking and dragging the middle unfilled box will translate your line changing only the y -intercept. Clicking and dragging any one of the outside filled boxes will rotate your line changing only the slope.

Move your line until you feel it is the best line to fit your data and write down your equation below. Use \hat{y} to represent predicted fat.

9. Identify the slope your equation. Interpret the slope in context.

10. Identify the y -intercept of your equation. What does your y -intercept value mean? Does this make sense in the context of the data?

Now get two other peoples equations and write down in the lines below.

Person's name_____ their equation_____

Person's name_____ their equation_____

How does your equation compare to the other two? Similar or different?

Check point here. Share some equations with class.

The Coefficient of Determination

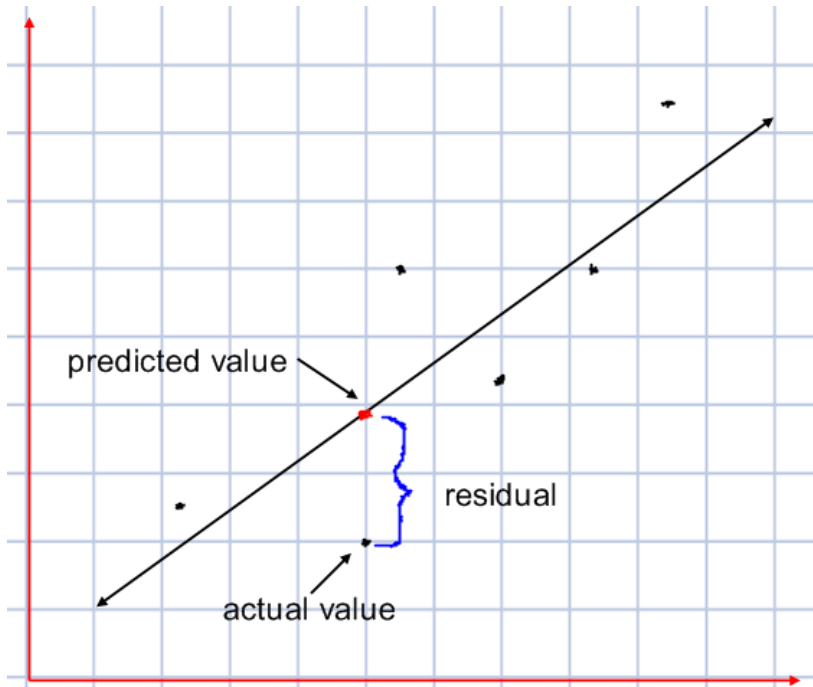
The square of r is called the coefficient of determination denoted by r^2 . The coefficient of determination measures the percentage of variation in the response variable that can be attributed to the variation in the predictor variable by the best-fit line. As the value gets closer to 1, the variation is better defined by the predictor variable and increasingly accurate predictions for the response variable can be assumed. The difference between 1 and r^2 , $(1 - r^2)$, represents the percentage of the variation in the response variable that is attributed to other variables besides the predictor variable.

10. Find r^2 by taking r and squaring it.
What is the value of r^2 ? Round to the nearest hundredth. _____.

Using the information from the above paragraph interpret what the coefficient of determination means.

Check point here. Share some interpretations with the class.

When we try to create a linear model by visually inspecting a graph, it is unlikely that two different people will generate the exact same line. If we have two or more different lines, how do we determine which is really best? There are different methods available and the purpose behind each method is to minimize the distance between the actual and predicted values. This distance is called a residual. Below is a picture of what a residual looks like.

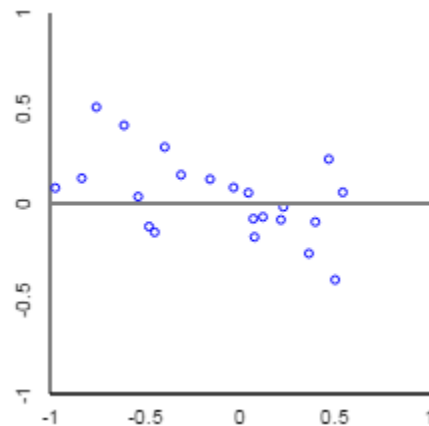
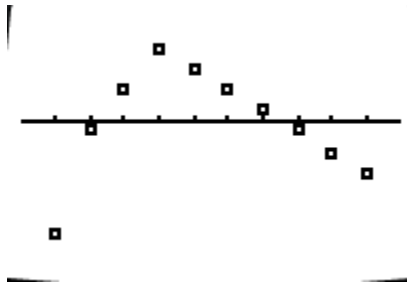



A residual is found by finding the difference between the actual and the predicted value determined by the regression line.

$$\text{residual} = \text{actual } y - \text{predicted } y$$

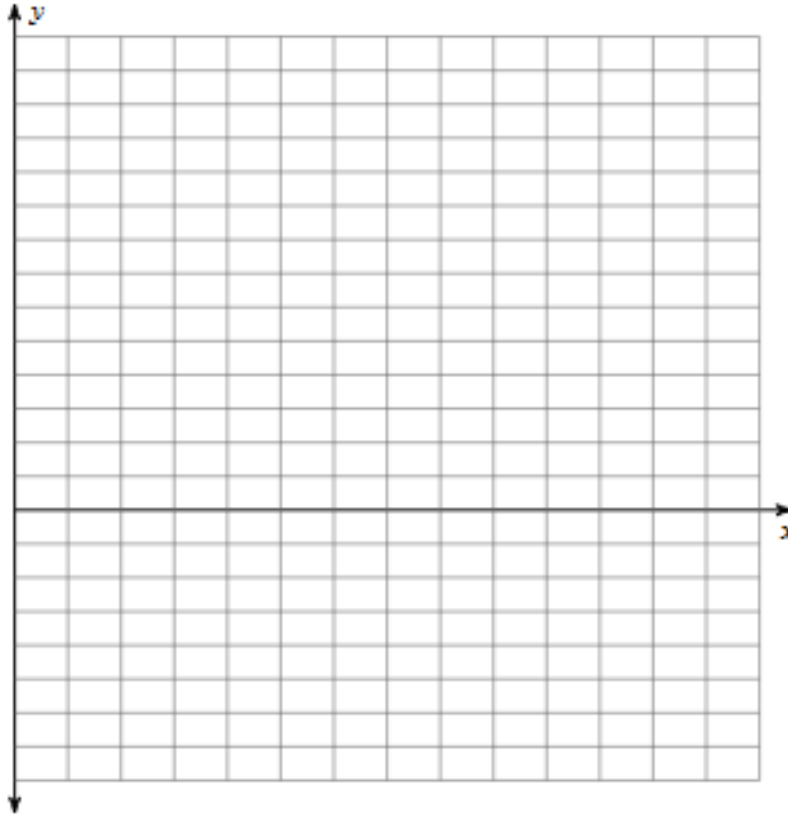
$$\text{residual} = y - \hat{y}$$

We can actually plot the residuals to help us determine whether a line is the best model for the two quantitative variables. Even if we have high r and r^2 values, it doesn't necessarily mean that a linear relationship is the best model between two variables. Creating a residual plot helps us determine whether a line is an appropriate model for the data. If there is no pattern to the plot, the residual plot does not curve. This indicates that a line is an appropriate model. If the residual plot shows a curved pattern, then the relationship between the variables is not linear. The residuals are calculated using the "Least Squares Regression" line. This is considered the best-fit line. The next lesson will explain in further detail why we call it the least squares regression. Below are examples of residual plots. Notice the residual on the left has a curve and the residual on the right shows no pattern.



Go to Core Math and click on  button. This will provide you what we call a residual plot.

11. Provide a sketch of the residual plot below. Please scale and label your axis.



a) Calculate the residual for the cheeseburger using the equation from #8. Show your work.

b) Calculate the residual for the Big N'Tasty using your equation from #8. Show your work.

Remember that a residuals value is based off the predicted value that comes from the regression line. You should notice that some residuals are positive and some are negative.

12. Using your two residuals from above, explain the meaning behind a positive and negative residual. (Hint: Over-estimation, under-estimation). What can you infer about the sum of your residuals?

13. Using the best-fit equation from #8, find the actual fat(g) of a sandwich whose residual is approximately 4.25 and carb(g) is 35. Interpret the meaning of this residual.