What Are Residuals
And
Why Do I Need Them?

Sharon Taylor
Georgia Southern University
taylors@georgiasouthern.edu

NCTM
Boston, MA
April 18, 2015

Below is a data set for homework averages and test grades for 10 students.

Homework Average	Test 1 Grade
56	42
75	75
92	90
91	97
78	74
77	59
69	74
80	83
80	87
55	45

One way to analyze this data is to use a scatterplot which is given for you on the next page.

The next procedure is adapted from "Spaghetti Regression" in Unit 4 of Coordinate Algebra from georgiastandards.org. Using spaghetti for regression lines is a common activity that allows for a hand-on way to involve students in choosing their line of best fit. However, it can also lead to a messy clean up.

Visualize what you think is a good line that fits the data. Use one of the colored markers or pencils at your table (along with a ruler if you like) and draw YOUR line of best fit on the scatterplot. Estimate an equation for your line.

y =

Now measure the vertical distance from each point on the scatterplot to YOUR regression line.

As you measure your vertical distances, you will notice that some will be above the line. Use a different color marker or pencil to mark these distances, which we will call positive.

Some distances will be below the line. Use yet another color marker or pencil to mark these distances, which we will call negative.

Add the distances and see what happens.

positive values = _____ negative values = _____ sum = _____

It is important to note variations from the original task. The task states students can measure horizontally, vertically, or perpendicular. However, residuals are defined by the vertical distances from a point to the line, which is what we are using here. Also, the task asks students to measure with spaghetti pieces and then tape all the pieces together to see which piece is shortest. Residual analysis is based on observing that the total (positive and negative) vertical distances of the "best fit" line is 0. In other words, be careful when looking for activities for residuals. There is a lot of incorrect information out there.

You should use the calculator to find the line of best fit. How close is your line to the calculator line?

Line of best fit: y =

You need to find the distances from the point to the line (y – value of each point minus the y – coordinate of your line).

Explore the residuals of **the** line of best fit instead of your line and see what happens.

x – value	y – value	predicted y – value (plug x into regression line)	difference (actual – predicted)

What happens when you add all the differences in the last column?

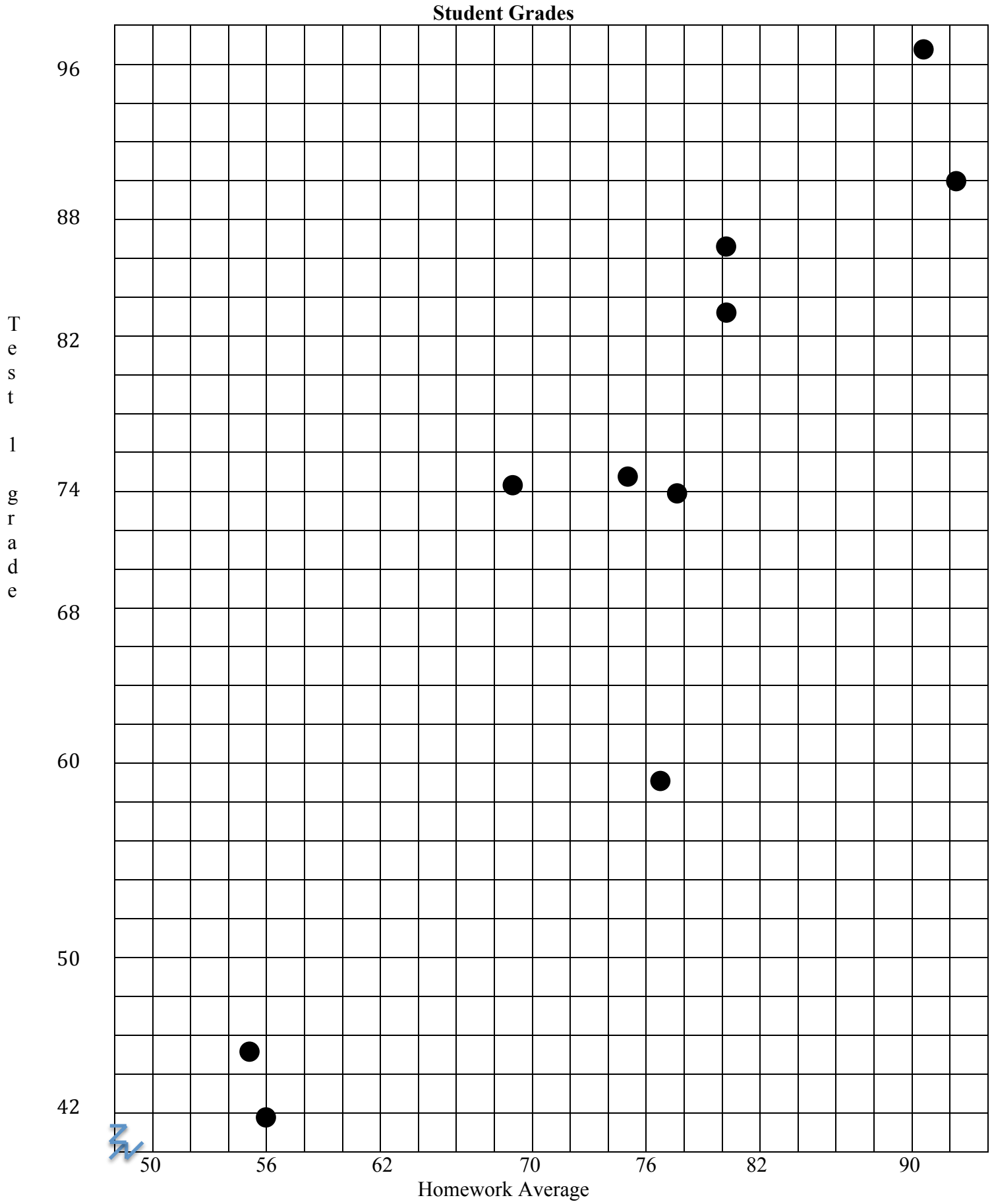
While it is important to go through a process of computing residuals by hand in order to understand what they are, it can be a tedious process. The calculator is very helpful in automating the process.

You can also have the calculator plot the distances so that you can look for a pattern.

The calculator should have also given you r and r^2 values. What do these values tell you about the goodness of fit for the best fit line? How does that compare with what you know about the residuals?

r =

r^2 =


Student Grades
96
88
82
74
68
60
50
42
Test 1 grade
50
56
62
70
76
82
90
Homework Average