Questions about the correlation coefficient?

Regression – fitting a line to data
Any line will not pass through all the points
Need a method for selecting a line of “best fit” to describe trends in the data
Again linear trends only

Method of least squares:
1. Look at the deviations of the points from the line in the Y-direction
   Residual = observed – predicted: \( y_i - \hat{y}_i \)
   One Residual for each observation – just like deviations

2. Compute the sum of the squared residuals \( \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \)

3. Choose the line that minimizes this sum of the squared residuals
   Least Squares Regression Line

Computing the least squares regression line.

Connections between regression, correlation and our descriptive statistics
Least Squares Equation: \( \hat{y}_i = a + bx_i \) for all values of \( x_i \)

\[
b = r \frac{S_y}{S_x}
\]

\[
a = \bar{y} - b\bar{x} = \bar{y} - r \frac{S_y}{S_x} \bar{x}
\]

Take note: using means and standard deviations of variables (what is the data model?)

Correlation and the interpretation of regression:
The square of the correlation coefficient (r) represents the fraction of the variation
in Y that is explained by the least squares regression using X as the explanatory variable.

Back to the In Class Exercise: Who came closest to the actual least squares line?

Mean x: 7.38    Mean y: 7.7
\( S_x=5.23 \)    \( S_y=4.11 \)
Slope: 0.64    Intercept: 2.98
Sum of the Squared Residuals: 14.79
Who came closest?
Who is furthest?

If we have time: Back to the NBA data
p:Data/MATH/StatM&M5/Ch02/ta2_001.MTP
Calculate the regression line for value as a function of income
Calculate the regression line for income as a function of value
Do they differ? Why? *What is different about the calculation?*
Does the r-squared value change between the two lines? Why or why not?

Reminders:
  a. Be done reading 2.4 for Wednesday. Problem session Friday.
  b. Homework #3 is ready for your perusal.