Statistical Reasoning
Least-Squares Regression Lines

Name: key Date: Class:

Coefficient of Determination (r^2)

The coefficient of determination (the square of the correlation coefficient, represents the fraction of variation in y value that can be explained by the reg. line

- The coefficient of determination r^2, is useful because it gives the % proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph.
- The coefficient of determination r^2 is the ratio of explained variation to the total variation.
- The coefficient of determination r^2 is such that $0 \leq r^2 \leq 1$, and denotes the strength of the linear association between x and y.
- The coefficient of determination r^2 represents the percent of the data that is the closest to the line of best fit. For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation). The other 15% of the total variation in y remains unexplained.
- The coefficient of determination r^2 is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.

Consider the relationship between messages sent versus messages received. It is not an exact 1 to 1 ratio, but there is a relationship of causation.

- Peoples' messages received (y) vary. They vary for many reasons, among them messages sent (x), day of the week, age, how long they have been texting, phone plan, and so on.
- Consider messages sent, and suppose the correlation between messages received and messages sent is r = 0.90 and r^2 = 0.81.
- Then differences in messages sent "explains" 81 % of the variability in messages received. The remaining 19 % of variability in messages received is attributable to other factors.
- In a sense, then, r^2 gives us an indication of how much students' messages sent tells us about differences in their messages received.
- In general, the higher the value of r^2, the more of the variability is explained, giving us a better tool for making estimates or predictions.