Day 22

Describing R2 in context: All Student Bistro sells plain pizzas for $8 and adding $1.50 per topping.
Create a scatterplot graphing cost of pizza against the number of toppings.
Describe the association:

The correlation, r = _____ so R^2 = _____
Now label the following statements as true or false.
   a) The number of toppings explains 100% of the cost of a pizza.
   b) Differences in the number of toppings explains 100% of the variation in the cost of pizzas.

Review A 21 (on OH)
We should always use meaningful variable names (in lieu of x and y) when writing regression lines.
$$\text{fat} = 8.7 + 0.61 \square \text{protein}$$

IPS Website: Correlation and Regression

A high R^2	Does not demonstrate the appropriateness of the regression.

So how do we check appropriateness?

3 conditions needed for Linear Regression Models: /* same as correlation */	1. Quantitative Variables 2. Straight Enough – check original scatterplot & residual scatterplot 3. Outlier (clusters) –points with large residuals and/or high leverage

But sometimes it is difficult to determine form or other problems with a scatterplot.
Fortunately not only does our line of best fit provide a convenient model for the association it also provides an excellent method for checking its appropriateness.

IPS Website: Two-variable Statistical Calculator

Looking at a _____ is a good way to check the Straight Enough Condition. It should be _____	a scatterplot of the residuals vs. the *x*-values. (appropriateness) If we extract all the form, Then boring: uniform scatter with no direction, shape, or outliers.
The _____ is the key to assessing how well the model fits.	variation in the residuals
Standard deviation of the residuals, s_e	Gives a measure of how much the points spread around the regression line.
$1 - R^2$	The fraction of the original variation left in the residuals. (The percentage of variability not explained by the regression line.)
Extrapolations	Dubious predictions of *y*-values based on *x*-values outside the range of the original data.

Let's use the calculator: Equation for Line of Best Fit, Draw Line, Residual Plots.
Fortunately we already know how to calculate the equation of the regression line, and R^2. It is done for us when we calculated correlation back on Day 19. However if we want to graph the regression line and make residual plots we have need some modification and additional steps (in **bold** below).