Day 20

Last week we learned to create scatterplots **to assess potential associations between two quantitative variables. We described the association by telling about** form, direction, strength, unusual features**. Then we learned how to calculate** correlation **to describe the strength of the association between two variables as long as these conditions are met:** quantitative, linear, no outliers

Remember, correlation measures the strength of association between x and y by measuring how closely the points of a scatterplot hug a line that goes through all the points that are an equal number of standard deviations away from the mean for both variables. Today we seek to figure out:

How can we use the correlation coefficient **to estimate the y value for a given x value?**

We call the estimate made from a model the predicted value, and write it as $\hat{y}$ (y-hat) to distinguish it from the true value of y.

If there is a perfect correlation (that is, $r = -1$ or $r = 1$) then we can perfectly predict y from x.

- r = 1 If x goes up 1 SD from average, then y goes up 1 SD from average
  All points lie on the SD line.
- r = .8 If x increases 1 SD, then y increases only 0.8 SD on the average
- r = .5 If x increases 1 SD, then y increases only 0.5 SD on the average
- r = -.3 If x increases 1 SD, then y decreases only 0.3 SD on the average
- r = 0 If x increases 1 SD, then y decreases only 0 SD on the average

Regression (explanatory) estimates the average value for the response variable (y) corresponding to each value of the variable (x).

1 SD increase in x means only an r SD increase in y.

Example 1: A large class took two exams with the following results:

$\bar{x}_{exam1} = 80$ $s_{exam1} = 10$, $\bar{x}_{exam2} = 70$ $s_{exam2} = 15$, $r = .6$

a) Suppose a student is picked at random. What would be your best guess for what he got on exam 2?

Don't know anything → guess any number

b) Now suppose you're told that the student got a 90 on Exam 1, what would be your best guess for his score on Exam 2?

EXAM 1: $z\text{-score} = \frac{90 - 80}{10}$

$z\text{-score} = 1$

r: $1 \cdot 0.6 = 0.6$

Exam 2: $0.6 = \frac{\hat{y} - 70}{15}$

$\hat{y} = 79$

c) A student got a 65 on Exam 1, but hasn't taken Exam 2 yet. What would you predict his score to be?

Exam 1: $z\text{-score} = \frac{65 - 80}{10}$

$z\text{-score} = -1.5$

r: $-1.5 \cdot 0.6 = -0.9$

Exam 2: $-0.9 = \frac{\hat{y} - 70}{15}$

$\hat{y} = 56.5$

Note: The estimated z-score is always closer to 0 than the given z-score unless perfect correlation

Regression to the mean	Because the correlation is always less than 1.0 in magnitude, each predicted $\hat{y}$ tends to be fewer SD from its mean than its corresponding x was from its mean.

A20

Complete the rest of the problems on the handouts until you see **Day 20 Stop**