

Math 58 / 58B - Introduction to (Bio)Statistics

your name here

due Friday, April 24, 2020

Homework 10

Assignment Summary (Goals)

- inference on linear model
- residual plots
- models with multiple variables

1. Correlation Properties¹

R: no technology for this problem

Suppose that you take a random sample of classes on campus and record the number of students enrolled in the class and the average evaluation score assigned by students to that teacher's effectiveness (on an A = 4, B = 3, ... scale).

- If the correlation coefficient turned out to be very close to zero, would you conclude that larger classes tend to have lower teaching evaluation averages? Explain.
- Suppose that the correlation coefficient turned out to be $r = -0.5$. Would you expect this to be more statistically significant if the sample size were $n = 5$ or if the sample size were $n = 50$, or would you expect sample size not to matter? Explain.
- Suppose that you also record whether the teacher was male or female. Would it make sense to calculate the correlation coefficient between class size and sex? Explain.

2. Least Squares Coefficients²

R: no technology for this problem

Re-consider the expressions for calculating least squares coefficients for the slope (calculation for slope is: $b_1 = r s_y / s_x$) and intercept ($b_0 = \bar{y} - b_1 \bar{x}$) of a regression line. Use these formulas to **explain what happens to the least squares line** in the following situations. [Hint: You may find it helpful to draw sketches as well as analyze these situations algebraically.]

For each part below, explain whether and/or how both b_0 and b_1 change.

- The mean value of the response variable increases, and all else remains the same. [Hint: Report what happens to slope and to the intercept.]
- The mean value of the explanatory variable increases, and all else remains the same.
- The standard deviation of the values of the response variable increases, and all else remains the same.
- The standard deviation of the values of the explanatory variable increases, and all else remains the same.
- The correlation coefficient between the two variables moves closer to zero, and all else remains the same.

3. Height and Foot Size³

¹From ISCAM, HW 5.32

²From ISCAM, HW 5.33

³From ISCAM, HW 5.35

R: just calculator

Reconsider the height and foot length data of Investigation 5.8 (HeightFoot.txt). In that investigation you used foot length (in centimeters) to predict height (in inches). Recall that some summary statistics are:

	Mean	SD	Correlation
Height	67.75 in	5.004 in	0.711
Foot length	28.50 cm	3.445 cm	

Now suppose that you want to use these data to construct a model for predicting a student's foot length from his/her height. (After all, most people can tell you their height off the top of their heads, but few can tell you the length of their right foot in centimeters.)

- (a) Use these summary statistics to determine the least squares regression line for predicting foot length from height.
- (b) Interpret the slope coefficient of this line in context.
- (c) Use the line to predict the foot length of a student who is 66 inches tall.
- (d) Identify the units of measurement (e.g., inches, centimeters, no units) on the slope coefficient, the intercept coefficient, and the correlation coefficient.
- (e) The least squares line for predicting height from foot length is $height = 38.3 + 1.03feet$. Solve this equation for foot length as a linear function of height.
- (f) Is the line in (e) the same as the line in (a)? Explain.