Math 10 Spring 2010 Quiz 4

Name: KEY April 30, 2010

(1) Suppose you have a list of entries to a repurposed-material design contest. Each uses some number of CDs and some quantity of fabric from curtains or sheets. The correlation coefficient for number of CDs and amount of fabric is -0.6.

3 (a) Suppose you take the points on the graph of averages (average fabric amount for each number of CDs) and find the correlation coefficient for that set. Do you expect it to increase or decrease from the original coefficient? Can you say for certain how it will change (if at all)?

If the data is homoscedastic, the value of r should decrease ($|r|$ increase) because extreme values are given more weight proportionally than they were previously. [the regression line may not move, but SD_x and SD_y have also changed, which accounts for that.]
If the data is heteroscedastic anything can happen, but as this is an example of ecological corr, we still expect $|r|$ to increase.

3 (b) Now suppose you compute the average number of CDs and amount of fabric used by entrants from each of the 20 colleges that participated, and find the correlation coefficient for that 20-element data set. Do you expect it to increase or decrease from the original coefficient? Can you say for certain how it will change (if at all)?

This is an ecological correlation so we expect r to decrease ($|r|$ to increase) though it need not.

(2) The amount of two chemicals in the bloodstream of fish in a certain population may be summarized as follows.

chemical A: average 0.02 ppm; SD 0.001 ppm
chemical B: average 0.05 ppm; SD 0.01 ppm
$r = 0.75$

5 (a) If you have measured the amount of chemical A in a given fish's bloodstream at 0.019 ppm, how much of chemical B do you expect to find?

method 1: regression line ← better if you want to predict B for multiple values of A, otherwise equivalent to:

$$y - .05 = \frac{.75 \cdot .01}{.001}(x - .02)$$

(y is B because it is what we want to predict)

for $x = .019$

$$y = .05 + \frac{.75 \cdot .01}{.001}(-.001)$$

$= .0425$ ppm

method 2: standard units

$.019 - .02 = -.001 = 1$ SD below the mean.

B should be $.75 \cdot 1 = .75$ SDs below the mean.

$.05 - (.75 \cdot .01) = .0425$

3 (b) The researchers dub fish with at least 0.022 ppm of A in their bloodstream "A-enhanced", and with at least 0.07 ppm of B in their bloodstream as "B-enhanced". They observe that most A-enhanced fish are not B-enhanced, and vice-versa, and postulate some regulation method in the fish's biology inhibits simultaneously high levels of A and B. Does the data support this hypothesis? Why or why not?

The researchers have chosen values 2 SDs above the mean for their definitions, and since r is .75, not 1, the regression effect says their observation is inevitable. There could be a biological mechanism which makes extremely high levels of chemical A less likely to coexist with extremely high levels of chemical B (contributing to the value of r being <1) but the data do not tell you this. They have committed the regression fallacy.