KNR 445: Fiendish correlation practice (but good practice for the quiz!)

1. Guess the correlations:

   i) For this one, the answers are, I think, quite clear. As the plots go from left to right, the correlation becomes increasingly large and positive. So the correct answer is:
   - Plot A: \( r = 0.09 \)
   - Plot B: \( r = 0.34 \)
   - Plot C: \( r = 0.53 \)
   - Plot D: \( r = 0.92 \)

   ii) This is less clear. Plots A, B, and C are quite similar to each other. So I’m not too bothered about getting this spot on, but I’d opt for these answers:
   - Plot A: \( r = -0.53 \)
   - Plot B: \( r = -0.61 \)
   - Plot C: \( r = -0.44 \)
   - Plot D: \( r = 0.83 \)

2. Subtle correlation points (remember, the important concept is covariation)

   i) Estimate the correlation in this diagram. Explain the reasoning behind your answer.

   The basic idea here is to notice that if variation in X occurs without covariation in Y (or vice versa, ie the line is either perfectly horizontal or vertical), the correlation will be zero, as the top of the correlation equation will be zero. However, if as it appears here there is some small amount of covariation in Y and X, so that there is some slope, the correlation could be very close to 1 (as the ellipse fits tightly to a straight line drawn through the points)
ii) Explain how (and why) the correlations would change between each of the following left-right pairs of diagrams

Answer:
The correlation would grow, as you are adding in large variation in X that also covaries with Y.

Answer:
If the point is at the mean of Y, then r would shrink (adding in variation in X without covariation in Y – enlarging bottom of equation). If not at the mean of Y, it could grow – adding variation in X with covariation in Y.

Answer:
If the point is at the mean of X and Y, the effect on r will be negligible. The point neither adds to the SD of X, the SD of Y, nor the covariation of X and Y.
3. Some other issues:
   i) What effect would it have on a correlation coefficient if both (or even one) variable(s) had a poor range (that is, the people being measured were all rather similar on one or other of the measures being used?)

   The r would be smaller. As the range on a variable grows, the chance of observing covariation with the other variable grows (the chance of its covariation outweighing random variation grows). See this applet for a demo: http://www.ruf.rice.edu/~lane/stat_sim/comp_r/Contents.html (animate the SD of X to see the effect of restricted range of X on the r with Y).

   ii) What effect would it have on a correlation coefficient if you uniformly increased the value of every value for one variable (for instance, measured weight in ounces instead of pounds when correlating with height in inches)?

   None at all.

   iii) Which is the strongest association in the following correlation matrix?

   
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.0</td>
<td>.72</td>
<td>.48</td>
<td>-.43</td>
</tr>
<tr>
<td>B</td>
<td>.72</td>
<td>1.0</td>
<td>.09</td>
<td>.19</td>
</tr>
<tr>
<td>C</td>
<td>.48</td>
<td>.09</td>
<td>1.0</td>
<td>-.15</td>
</tr>
<tr>
<td>D</td>
<td>-.43</td>
<td>.19</td>
<td>-.15</td>
<td>1.0</td>
</tr>
</tbody>
</table>

   0.72

   iv) Which correlation coefficient is more important? Why? (And circle the coefficient of determination in each example)

   a. \( r = .6, R^2 = .36, p > .05 \)
   b. \( r = .5, R^2 = .25, p < .05 \)

   Well, you could argue this. The first has the stronger effect, as the \( R^2 \) is larger. But the second is significant, where the first is not. So the second is the more reliable effect. In reality, this is likely to be an issue of sample size. The first \( r \) requires a larger sample size (for more power – see later in course).

   v) If you increase \( \alpha \) (say, from 0.05 to 0.2), what happens to your chances of making a type I error?

   a. increase
   b. decrease
   c. stay the same
   d. depends on the data

   \( \alpha \) is your chance of making a type I error – it is the chance of getting a significant result when you should in fact not get one. So as \( \alpha \) increases, so does type I error rate.