1. A school suspects that there is a relationship between MBA students' GMAT scores and their grade point averages (GPAs) upon graduation. They take a simple random sample of 5 students and record their GMAT score and GPA:

GMAT score	660	580	480	710	600
GPA	3.7	3.0	3.2	4.0	3.5

   (a) How much better on average are the GPAs of students that received a GMAT score of 700 than those of students who received a GMAT score of 600?

   In this case, we have GMAT score as the predictor (x) variable and GPA as the response (y) variable. Our first step is to compute how many SDs of change we are making in the predictor variable. The predictor variable is GMAT score. Since the change is $700 - 600 = 100$*, and the SD for GMAT score was 87.06, the change is* $\frac{100}{87.06} = 1.15SDs$ *of GMAT score. Next, we multiply the number of the SDs that the predictor changed by the correlation. This is how many SDs the response will change. So, GPA will go up by* $1.15 * 0.82 = 0.94SDs$*. Finally, we multiply the number of SDs that the response changed by the SD for the response. Since the SD of GPA was 0.396, GPA will increase by* $0.396 * 0.94 = 0.37$*. I also think it helps to draw this out. How far are we moving in the X axis, then change in the Y direction by corresponding amount, correlation times number of SDs of change in X.*

   (b) Write a linear relationship between GMAT score (X) and GPA (y), so you're able to predict GPA as your response (y). Interpret the slope and y-intercept in context.

   slope $= r * \frac{SD(Y)}{SD(X)} = 0.824 * 0.39687.06 = 0.0037$*. For every one point increase in GMAT score, the GPA is predicted to increase on average by 0.0037 units. Equation of line:*$\hat{y} = 0.0037 * x + b*$*, and we know it goes through the average of both variables:* $3.48 = 0.0037 * 606 + b$*, so* $b = 3.48 - 0.0037 * 606 = 1.238*$*. The predicted GPA for someone with GMAT score of 0 is 1.238 (Avg GPA for GMAT score of 0). Equation of line:*$\hat{y} = 0.0037 * X + 1.238$*. The key to remebere here is that we have a forumla for a slope and we know the line must go through the average of X and the Average of Y. With this information we can find the one unique, correct line.*

   (c) Suppose a student had a GMAT score of 640. What would you predict their GPA to be?

   $\hat{y} = 0.0037 * 640 + 1.238 = 3.606$

   (d) Suppose you were asked to predict the GPA of a student with a GMAT score of 790. Why is this a bad idea?

   The largest GMAT score in the available data was 710. There's no guarantee that the relationship will still be linear and have the same slope or intercept past this value. So we should not predict for a GMAT score of 790 because that would be extrapolation. Extrapolation has been the down fall of many. A great example is that by 2050, women are predicted to be running marathons in under a minute. In the 70's and 80's women were dramatically improving marathon times. People built models assuming this is a linear relationship and will continue forever. That was a pretty obvious mistake but people still made it.

2. A study of department chairperson ratings and student ratings of the performance of high school statistics teachers reports a correlation of $r = 1.20$ between the two ratings. From this information we can conclude that:

   (A) Chairpersons and students tend to agree on who is a good teacher.

   (B) Chairpersons and students tend to disagree on who is a good teacher.

   (C) There is little relationship between chairperson and student ratings of teachers.

(D) There is strong association between chairperson and student ratings of teachers, but it would be incorrect to infer causation. D is close to the right idea but again r cant be less then -1.

(E) Some sort of mistake has been made. Correlations must be between -1 and $+1$.

3. For the following research question, identify at least one analysis that would be appropriate for the situation.

   What is the relationship between the amount of alcohol consumed (in ounces) and the level of coordination (on a scale from 1 to 10) measured on a set of individuals? We have two continuous measurements, each taken from the same experimental units. We can calculate the correlation coefficient between the two variables. We could also consider writing a linear relationship between them (if there is a linear pattern to the points (amt, coordination)