KRN 445: Some regression practice

1) Say you have a regression equation of $Y' = 85 + 6X$. $Y'$ is estimated IQ, $X$ is years in education. $R^2$ is .25, and the standard error of the estimate is 10 IQ points.
   a) What is the slope, and what is the intercept?

   The slope is 85 IQ points, the intercept is 6 IQ points.
   b) What would be the predicted IQ for someone with 10 years of education?

   From $Y' = 85 + 6 (10)$, $Y' = 85 + 60 = \boxed{145}$

   c) Say you found a person with an IQ of 120 who had received 10 years of education. What would be the residual for this person?

   For 10 years of ed., $Y' = 85 + 60 = 145$, as before. An actual $Y$ of 120 would then be 25 IQ points below the line, so the residual would be $\boxed{-25}$.

   d) How much higher of an IQ does the model predict for someone who has 8 years of education than someone who has 5?

   The slope is 6 IQ points per year of education. So, for three years, an increase of $18$ IQ pts. would be expected.

   e) What range of IQ’s might we expect to see 95% of the time for 5 years in education?

   This would be obtained from the $Y' +/ - 2 \text{SE}_E$. So, as $Y' = 85 + 30 = 115$, and $\text{SE}_E = 10$, the range would be $115 +/ - 20 = \boxed{95 \text{ to } 135}$.

   f) How about a range of IQ’s expected 68% of the time for 10 years of education?

   This would be obtained from the $Y' +/ - 1 \text{SE}_E$. So, as $Y' = 85 + 60 = 145$, and $\text{SE}_E = 10$, the range would be $145 +/ - 10 = \boxed{135 \text{ to } 155}$.

2) If the sample upon whom the previous equation is based had only a small range of years of education, what effect would this have on the equation’s ability to predict effects of larger periods of education, and why?

   It would be reduced, as the proportion of explained variation would be reduced relative to the unexplained variation (as regression function is always best at predicting values within the measured range of variation).

3) What is the relationship between the coefficient of determination and the standard error? (See slides)

4) What is the conceptual similarity between the SD for a normal distribution, and the SE for regression? (see slides)