36-617: Applied Linear Models
Fall 2018
HW02 – Solution to Problem 4

4. [Gelman & Hill (2007), Ch 3, #2] Suppose that, for a certain population, we can predict log earnings from log height as follows:

   - A person who is 66 inches tall is predicted to have earnings of $30,000.
   - Every increase of 1% in height corresponds to a predicted increase of 0.8% in earnings.
   - The earnings of approximately 95% of people fall within a factor of 1.1 of predicted values.

   (a) Give the equation of the regression line and the residual standard deviation of the regression.
   (b) Suppose the standard deviation of log heights is 5% in this population. What, then, is the R^2 of the regression model described here?

Solution

This problem is really not about the regression analysis problems we have been discussing in class, but rather it is about the conditional distribution of Y given X when X and Y have a joint bivariate normal distribution,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} \right)$$

For example, the basic facts needed about the bivariate normal distribution (alas, with X and Y reversed) can be found here: `http://athenasc.com/Bivariate-Normal.pdf`. It turns out that the conditional distribution of Y given X is

$$Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2_{Y|X})$$

where

$$\begin{aligned} \beta_1 &= \rho \frac{\sigma_Y}{\sigma_X} & (1) \\ \beta_0 &= \mu_Y - \beta_1 \mu_X & (2) \\ \sigma^2_{Y|X} &= (1-\rho^2)\sigma_Y^2 & (3) \end{aligned}$$

These formulas are similar to but not identical with the formulas we have learned from regression analysis, largely because of the close relationship of the normal distribution with least-squares estimation.

Now let's apply these ideas with

$$\begin{aligned} Y &= \log(\textit{earnings}) \\ X &= \log(\textit{height}) \end{aligned}$$

From what is given in the problem, we know

$$\begin{aligned} \rho \frac{\sigma_Y}{\sigma_X} = \beta_1 &= 0.8 & (4) \\ \beta_0 &= \log(30000) - (0.8)\log(66) = 6.96 & (5) \\ 2\sqrt{(1-\rho^2)\sigma_Y^2} = \sigma_{Y|X} &= 0.1 & (6) \\ \sigma_X &= 0.05 & (7) \end{aligned}$$

Equations (4), (5) and (6) give the answers to 4(a). For 4(b) we square equations (4) and (6), make use of equation (7), and rearrange, giving

$$\begin{aligned} \rho^2\sigma_Y^2 &= (0.8)^2(0.05)^2 = 0.0016 \\ (1-\rho^2)\sigma_y^2 &= (0.01)^2 \times 2^2 = 0.0004 \end{aligned}$$

and hence

$$\begin{aligned} \sigma_Y^2 &= 0.0016 + 0.0004 = 0.0020 \\ \rho^2 &= 0.0016/0.0020 = 0.8; \end{aligned}$$

this gives the answer for 4(b).