STAT 51200 – FALL 2022
Applied Regression Analysis

Homework #04

1. Finish reading Chapter 2 in the text.
2. Suppose (X, Y) are bivariate Normal, and therefore have the joint density function

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times D(x,y), \qquad \forall (x,\, y) \in \mathbb{R}$$

where

$$D(x,y) \equiv \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}.$$

   a) What is the conditional distribution of $Y|X = x$?
   b) Given a sample of n *i.i.d.* pairs $(x_i, y_i), i = 1, 2, \cdots, n$, from the above bivariate Normal distribution, express the simple linear regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2)$, *i.i.d.* by using the five parameters, $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$, that is, expressing β_0, β_1 and σ^2 in terms of $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$.
3. Do Problems 2.50, 2.51, 2.52, 2.55, 2.58
4. Six observations on Y are to be taken when X values are $x = 4, 8, 10, 12, 16, 20$, respectively. The true regression *line* is $E(Y) = 20 + 4X$ and the error terms ϵ_i are known to be independent $\mathcal{N}(0, 5^2)$ random variable. Use **R** for the following "tasks".
   a) Generate $n = 6$ $\mathcal{N}(0, 5^2)$ random error terms (`rnorm(6, 0, 5)`), and use them to simulate the corresponding $n = 6$ values of Y $(y_1, y_2, \ldots, y_6)$. Obtain the corresponding least squares estimates b_0 and b_1 of β_0 and β_1. Also, obtain the Least Squares estimated of $E(Y_h)$ when $X_h = 14$ and construct a 95% Confidence Interval for it.
   b) Repeat part a) $N = 2000$ times, simulating new sample values each time.
   c) Make a histogram of the $N = 2000$ estimates b_1 you obtained above and calculate th emean and standard deviation of these estimates. Are your results consistent with the theoretical expectations?
   d) What proportion of your $N = 2000$ calculated confidence intervals for $E(Y_h)$ (when $X_h = 14$) you found to include the *"true"* value of it? Are your results consistent with the theoretical expectations?