2. (Fox 5.5) Using calculus, derive the normal equations

$$\text{OLS} = \underset{(\beta_0, \beta_1, \beta_2 \cdots \beta_k)}{\arg\min} \quad \sum_{i=1}^{n}\left[Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_k X_{ik})\right]^2$$

$\beta_0:$

$$\frac{d}{d\beta_0} = -2\sum_{i=1}^{n}\left[Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})\right]$$

Setting to 0, we have:

$$\sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum X_{i1} + \cdots + \beta_k \sum X_{ik} \quad ------- \text{①}$$

$\beta_j \in \{\beta_1, \dots, \beta_k\}$

$$\frac{d}{d\beta_j} = -2\sum_{i=1}^{n} X_{ij}\left[Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})\right]$$

Setting to 0, we have,

$$\sum_{i=1}^{n} X_{ij} Y_i = \sum_{i=1}^{n} X_{ij}(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

i.e.
$$\sum_{i=1}^{n} X_{ij} Y_i = \beta_0 \sum_{i=1}^{n} X_{ij} + \beta_1 \sum_{i=1}^{n} X_{i1} X_{ij} + \cdots + \beta_j \sum X_{ij}^2 + \cdots \beta_k \sum X_{ij} X_{ik} \quad ---\text{②}$$

normal equations.

3. (Fox 6.5) Maximum likelihood estimation of the simple-regression model

Note that the joint probability density for the Y_is is a product of their marginals under independence. The marginals are:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma_\varepsilon^2}\right]$$

Therefore the log-likelihood is:

$$l(\alpha, \beta, \sigma_\varepsilon^2) = \log\left(\prod p(y_i)\right)$$

$$l(\alpha, \beta, \sigma_\varepsilon^2) = \sum_{i=1}^{n}\left(-\frac{1}{2}\log(2\pi\sigma_\varepsilon^2) + \left[-\frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma_\varepsilon^2}\right]\right)$$

which $\alpha, \beta, \sigma_\varepsilon^2$

Taking the derivative with respect to α yields:

let the maxi value of $\prod p(y_i)$

$$\frac{\partial}{\partial \alpha} l(\alpha, \beta, \sigma_\varepsilon^2) = \sum_{i=1}^{n}\left[\frac{2(y_i - (\alpha + \beta x_i))}{2\sigma_\varepsilon^2}\right] = 0$$

$$\implies \sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$$

$$\implies \hat{\alpha} = \bar{y} - \beta\bar{x}$$

② Next, taking the derivative with respect to β yields:

$$\frac{\partial}{\partial \beta} l(\alpha, \beta, \sigma_\epsilon^2) = \sum_{i=1}^{n} \left[\frac{2x_i(y_i - (\alpha + \beta x_i))}{2\sigma_\epsilon^2} \right] = 0$$

$$\implies \sum_{i=1}^{n} x_i(y_i - \alpha - \beta x_i) = 0$$

Using $\hat{\alpha} = \bar{y} - \beta\bar{x}$ gives:

$$\sum_{i=1}^{n} x_i(y_i - (\bar{y} - \beta\bar{x}) - \beta x_i) = 0$$

$$\implies \sum_{i=1}^{n} (x_i y_i - x_i \bar{y}) = \beta \sum_{i=1}^{n} (x_i^2 - x_i \bar{x})$$

$$\implies \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i y_i - x_i \bar{y})}{\sum_{i=1}^{n}(x_i^2 - x_i \bar{x})} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

③ Finally, taking the derivative with respect to σ_ϵ^2 yields:

$$\frac{\partial}{\partial \sigma_\epsilon^2} l(\alpha, \beta, \sigma_\epsilon^2) = \sum_{i=1}^{n} \frac{1}{2\sigma_\epsilon^2} - \frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma_\epsilon^4} = 0$$

$$\implies \hat{\sigma_\epsilon^2} = \sum_{i=1}^{n} \frac{(y_i - (\alpha + \beta x_i))^2}{n}$$

(the cost of including irrelevant explanatory variables)

4. (Fox 6.9) Suppose that the "true" model generating a set of data is $Y = \alpha + \beta X_1 + \varepsilon$. A researcher fits the model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, which includes the explanatory variable X_2 - that is, the true value of β_2 is 0. Had the researcher fit the (correct) simple regression model, the variance of B_1 would have been (using the simple regression formula) $V(B_1) = \sigma_\varepsilon^2 / \sum (X_{i1} - \bar{X}_1)^2$.

a) Is the model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ wrong? In other words, is B_1 for this model a biased estimator of β_1?

b) What is the variance of B_1 in this model?

c) Considering (a) and (b), what is the cost of including the irrelevant explanatory variable X_2? How does this compare to the cost of *failing* to include a *relevant* explanatory variable?

Problem 4

(a) The model is not wrong. We simply have prior knowledge of the value for β_2. For this reason, the least squares estimate for the coefficient of X_1 is unbiased.

(b) The cost is that the variance for the estimated coefficient will be inflated by $1/(1 - r_{12}^2)$.

On the other hand, if we fail to include a variable, then the estimate for β_1 is biased. According to equation (6.8) in Fox, the bias is $\beta_2 \sigma_{12} / \sigma_1^2$. This bias is a problem if $\beta_2 \neq 0$ and X_1 and X_2 are correlated (so $\sigma_{12} \neq 0$).