1 The Simple Linear Regression Model

The linear regression model with one predictor variable is given by the below.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

1. Interpret β_0. What are the units of β_0?
   (a) If $X = 0$ is not within the range of the observed X values, is β_0 extrapolation or interpolation?
   (b) Give an example when $X = 0$ is meaningful from your work.
2. Interpret β_1. What are the units of β_1?
3. How does ordinary least squares (OLS) choose the parameter estimates for the fitted/estimated/predicted model $\widehat{y} = b_0 + b_1 x$? If you use any other values other than the estimates provided by OLS, what will happen to $SSE = \sum(y - \widehat{y})^2$?
4. What population parameter does $\widehat{y}$ estimate?
5. If a model overpredicts the value of y for a given x, is the residual positive or negative, and is the point above or below the line $\widehat{y}$?

2 The Multiple Linear Regression Model

The linear regression model with one predictor variable is given by

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$$

and the estimated model is given by

$$\widehat{y} = b_0 + b_1 x_1 + \ldots + b_k x_k$$

1. How does the interpretation of β_0 differ when there are multiple predictors?
2. How does the interpretation of β_i $(i = 1, \ldots, k)$ differ when there are multiple predictors?
3. What do we assume about the distribution of ε?
   (a) As a result, what do we assume about the distribution of Y and why?
   (b) How do you check this before fitting the model?
   (c) How do you check this after fitting the model?

4. What do we assume of about the variation of ε?
   (a) What does this mean in less technical terms?
   (b) How do you check this before fitting the model?
   (c) How do you check this after fitting the model?
5. What do we assume about the relationships between Y and the X_i in the model?
   (a) Give two examples of when the zero mean condition is not satisfied.
   (b) How do you check this before fitting the model?
   (c) How do you check this after fitting the model?
6. What do we assume about the relationships between $X_1, \ldots X_k$?
   (a) How do you check this before fitting the model?
   (b) How do you check this after fitting the model?
7. What do we assume about the the autocorrelation between the ε_i?
   (a) How do you check this after fitting the model?
   (b) Is the autocorrelation of the ε_i the same as the Y_i?
8. In regression, what is the difference between endogineity and exogeneity?
   (a) For a given fitted model, how might you be able to detect endogineity by examining the model coefficients?
   (b) If the above is detected, violation of what other condition could cause the same problem? How can you tell the difference?
   (c) Suppose the nature of the endogienity is omitted variable bias such that $COV[X_i, \varepsilon] \neq 0$ for some X_i in the model and the effects of an external X_j are accounted for by ε. What does this suggest about X_i and X_j? Would problem might introducing X_j cause?
   (d) Suppose there is a non-linear relationship between the X_i and ε. What might this suggest about the underlying model formulation?
9. What do we require to establish causality?