Solution to Tutorial 6

1. For each of the following regression models, indicate whether it is a general linear regression model. If not, state whether it can be expressed in the form of a linear regression model after some suitable transformation

$$
\begin{array}{ll}
a. & Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \log X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i \\
b. & Y_i = \varepsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2), \quad with\ \varepsilon_i > 0 \\
c. & Y_i = \beta_0 \log(\beta_1 X_{i1}) + \varepsilon_i \\
d. & Y_i = \log(\beta_1 X_{i1}) + \beta_2 \log X_{i2} + \varepsilon_i \\
e. & Y_i = [1 + \exp(\beta_0 + \beta_1 X_{i1} + \varepsilon_i]^{-1}
\end{array}
$$

No of these are linear models, but a, b, d e can be transformed to linear regression models

2. Consider the multiple linear regression models

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \ i = 1, ..., n$$

where ε_i are uncorrelated with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2$ state the least square criterion and derive the least squares estimators for β_1 and β_2. Let

$$Q(b_1, b_2) = \sum_{i=1}^n \{Y_i - b_1 X_{i1} - b_2 X_{i2}\}^2$$

by calculus, we have

$$
\begin{aligned}
\frac{dQ(b_1, b_2)}{db_1} &= -2 \sum_{i=1}^n \{Y_i - b_1 X_{i1} - b_2 X_{i2}\} X_{i1} \\
\frac{dQ(b_1, b_2)}{db_2} &= -2 \sum_{i=1}^n \{Y_i - b_1 X_{i1} - b_2 X_{i2}\} X_{i2}
\end{aligned}
$$

The normal equations are

$$
\begin{aligned}
\sum_{i=1}^n X_{i1}^2 b_1 + \sum_{i=1}^n X_{i1} X_{i2} b_2 &= \sum_{i=1}^n X_{i1} Y_i \\
\sum_{i=1}^n X_{i1} X_{i2} b_1 + \sum_{i=1}^n X_{i2}^2 b_2 &= \sum_{i=1}^n X_{i2} Y_i
\end{aligned}
$$

Solving it we have

$$
\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1} X_{i2} \\ \sum_{i=1}^n X_{i1} X_{i2} & \sum_{i=1}^n X_{i2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n X_{i1} Y_i \\ \sum_{i=1}^n X_{i2} Y_i \end{pmatrix}
$$

3. Consider the multiple regression models

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \varepsilon_i, \ i = 1, ..., n$$

where ε_i are uncorrelated with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2$. state the least square criterion and derive the least squares normal equations.

4. An analyst wanted to fit the regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \ i = 1, ..., n$$

by the least squares estimation when it is known that $\beta_2 = 4$. How can he implement the calculation in computer code?

The least squares become

$$\begin{aligned} Q &= \sum_{i=1}^{n} \{Y_i - b_0 - b_1 X_{i1} - 4X_{i2} - b_3 X_{i3}\}^2 \\ &= \sum_{i=1}^{n} \{(Y_i - 4X_{i2}) - b_0 - b_1 X_{i1} - b_3 X_{i3}\}^2 \end{aligned}$$

Thus it is equivalent to model

$$Z_i = \beta_0 + \beta_1 X_{i1} + \beta_3 X_{i3} + \varepsilon_i, \ i = 1, ..., n$$

where $Z_i = Y_i - 4X_{i2}$

5. For the multiple regression model, let r be the sample linear correlation coefficient between Y_i and $\hat{Y}_i$, and R^2 be the multiple determination of the model. Show that $R^2 = r^2$.

$$\begin{aligned} r &= \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})}{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y} + e_i)}{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \frac{\sum_{i=1}^{n}\{(\hat{Y}_i - \bar{Y})^2 + (\hat{Y}_i - \bar{Y})e_i\}}{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})e_i\}}{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \frac{\{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2\}^{1/2}}{\{\sum_{i=1}^{n}(Y_i - \bar{Y})^2\}^{1/2}} \\ &= \{R^2\}^{1/2} \end{aligned}$$

6. In a small-scale regression study, the following data were obtained

i:	1	2	3	4	5	6
X_{i1}	7	4	16	3	21	8
X_{i2}	33	41	7	49	5	31
Y_i	42	33	75	28	91	55

Assume that the multiple linear regression model with independent error terms is appropriate. Using matrix methods, obtain

(a) $\mathbf{b}$, (b) $\mathbf{e}$ (c) SR, (d) $\mathbf{s^2}(\mathbf{b})$, (e) $\hat{Y}$ when $X_1 = 10, X_2 = 30$, (f) $s^2(\hat{Y})$ when $X_1 = 10, X_2 = 30$

See **code**