Quiz 1- Linear Regression Analysis (Based on Lectures 1-14) Time: 1 Hour

1. In the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, with $\mathrm{E}(\varepsilon_i) = 3,\ \mathrm{E}(\varepsilon_i^2) = \sigma^2$, $i = 1, 2, ..., n$, the unbiased direct least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 respectively, are

(A) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{\beta}_1 = \dfrac{s_{xy}}{s_x^2}$

(B) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} - 3$, $\hat{\beta}_1 = \dfrac{s_{xy}}{s_x^2} - 3$

(C) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{\beta}_1 = \dfrac{s_{xy}}{s_x^2} - 3$

(D) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} - 3$, $\hat{\beta}_1 = \dfrac{s_{xy}}{s_x^2}$

where $s_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}),\ s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2,\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$

Answer: (C)

<u>Solution:</u> Minimizing $\sum_{i=1}^{n} \varepsilon_i^2$ with respect to β_0 and β_1 gives

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial S}{\partial \beta_0} = 0, \frac{\partial S}{\partial \beta_1} = 0 \Rightarrow \hat{\beta}_1 = \frac{s_{xy}}{s_x^2}, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Now express

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i y_i \text{ where } k_i = \frac{x_i - \bar{x}}{s_x^2}$$

$$\mathrm{E}(\hat{\beta}_1) = \sum_{i=1}^{n} k_i \mathrm{E}(\beta_0 + \beta_1 x_i + \varepsilon_i)$$
$$= 0 + \beta_1 + 3$$
$$\mathrm{E}(\hat{\beta}_1 - 3) = \beta_1.$$

So $\hat{\beta}_1 - 3 = \dfrac{s_{xy}}{s_x^2} - 3$ is an unbiased estimator of β_1.

Next

$$\begin{aligned}
\mathrm{E}(\hat{\beta}_0) &= \mathrm{E}(\overline{y} - \hat{\beta}_1 \overline{x}) \\
&= \mathrm{E}(\beta_0 + \beta_1 \overline{x} + \overline{\varepsilon} - \hat{\beta}_1 x) \\
&= \beta_0 + \beta_1 \overline{x} + 3 - \overline{x}(\beta_1 + 3) \\
&= \beta_0 .
\end{aligned}$$

So $\hat{\beta}_0 = \dfrac{s_{xy}}{s_x^2}$ is an unbiased estimator of β_0.

2. In the simple linear model $y_i = \beta_0 + \beta_i x_i + \varepsilon_i, \mathrm{E}(\varepsilon_i) = 0, \mathrm{E}(\varepsilon_i^2) = \sigma_i^2, i = 1, 2, ..., n,$ and assume that σ_i^2 is known. The best linear unbiased estimator of β_1 is

(A) $\dfrac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{1=1}^{n} (x_i - \overline{x})^2}$

(B) $\dfrac{\sum_{i=1}^{n} \sigma_i^2 (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} \sigma_i^2 (x_i - \overline{x})^2}$

(C) $\dfrac{\sum_{i=1}^{n} \dfrac{(x_i - \overline{x})(y_i - \overline{y})}{\sigma_i^2}}{\sum_{i=1}^{n} \dfrac{(x_i - \overline{x})^2}{\sigma_i^2}}$

(D) $\dfrac{\sum_{i=1}^{n} \left(\dfrac{x_i}{\sigma_i} - \sum_{i=1}^{n} \dfrac{x_i}{n\sigma_i} \right)\left(\dfrac{y_i}{\sigma_i} - \sum_{i=1}^{n} \dfrac{y_i}{n\sigma_i} \right)}{\sum_{i=1}^{n} \left(\dfrac{x_i}{\sigma_i} - \sum_{i=1}^{n} \dfrac{x_i}{n\sigma_i} \right)^2}$

Answer: (D)

Solution: The Gaurs Markoff theorem tells that the best linear unbiased estimator of β_1 is

$b = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$ when $\mathrm{E}(\varepsilon_i^2)$ is constant and independent of i. So transform the model

$y_i = \beta_0 + \beta_i x_i + \varepsilon_i, \mathrm{E}(\varepsilon_i) = 0, \mathrm{E}(\varepsilon_i^2) = \sigma_i^2, i = 1, 2, ..., n,$ as

$$\frac{y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \frac{\beta_1 x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$

or $y_1^* = \beta_0^* + \beta_1 x_i^* + \varepsilon_i^*$,

where $y_1^* = \dfrac{y_i}{\sigma_i},\ x_i^* = \dfrac{x_i}{\sigma_i},\ \varepsilon_i^* = \dfrac{\varepsilon_i}{\sigma_i}$. In the transformed model, we have $\mathrm{E}(\varepsilon_i^*) = 0, \mathrm{E}(\varepsilon_i^{*2}) = 1$.

Thus the best linear unbiased estimator of β_1 is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i^* - \overline{x}^*)(y_i^* - \overline{y}^*)}{\sum_{i=1}^{n}(x_i^* - \overline{x}^*)^2} = \frac{\sum_{i=1}^{n}\left(\frac{x_i}{\sigma_i} - \sum_{i=1}^{n}\frac{x_i}{n\sigma_i}\right)\left(\frac{y_i}{\sigma_i} - \sum_{i=1}^{n}\frac{y_i}{n\sigma_i}\right)}{\sum_{i=1}^{n}\left(\frac{x_i}{\sigma_i} - \sum_{i=1}^{n}\frac{x_i}{n\sigma_i}\right)^2}.$$

3. Consider the multiple linear regression model $y = X\beta + \varepsilon$ where y is a $n \times 1$ vector of observations on response variable, X is a $n \times K$ matrix of n observations on each of the K explanatory variables, β is a $K \times 1$ vector of regression coefficients and ε is a $n \times 1$ vector of random errors with $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{E}(\varepsilon\varepsilon') = \Omega$. The covariance matrix of the ordinary least squares estimator of β is

(A) $(X'X)^{-1}$

(B) $(X'X)^{-1}X'\Omega X(X'X)^{-1}$

(C) $(X'\Omega X)^{-1}X'X(X'\Omega X)^{-1}$

(D) $(X'X)^{-1}X'\Omega^{-1}X(X'X)^{-1}$

Answer: (B)

Solution: The ordinary least squares estimator of β is

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon).$$

The estimation error of $\hat{\beta}$ is

$$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$$

and the covariance matrix of $\hat{\beta}$ is

$$\begin{aligned} Cov(\hat{\beta}) &= \mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\ &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'\mathrm{X})^{-1}] \\ &= (X'X)^{-1}X'\mathrm{E}(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1}. \end{aligned}$$

4. The values of coefficient of determinations in the two multiple linear regression models are 0.4 and 0.7. Let y be the study variable, and X_i's $(i = 1, 2, 3)$ be the explanatory variables. These values of coefficient of determinations belongs to which of the two possible fitted models out of following four possible fitted models,

I. $y = 3 + 3X_1 + 3X_2 + 3X_3$

II. $y = 0.2 + 0.3X_1 + 0.4X_2 + 0.5X_3$

III. $y = -3X_1 - 3X_2 - 3X_3$

IV. $y = -0.3X_1 - 0.4X_2 - 0.5X_3$

(A) (I) and (II)

(B) (III) and (IV)

(C) (I) and (III)

(D) (II) and (IV)

Answer: (A)

Solution: The coefficient of determination is defined only when the intercept term is present in the linear model. Since only the linear models in (I) and (II) are having intercept terms, so the coefficient of determination can be defined only in these two possible models.

5. The observations on study (y) and explanatory variables X_i 's in a usual multiple linear regression model with four explanatory variables are standardized, i.e., every observation is considered as deviation from its mean and is divided by its standard deviation. Which of the following two models represent the possible fitted models with standardized observations:

I. $y = 1 + X_1 + X_2 + X_3 + X_4$

II. $y = \frac{X_1}{3} + \frac{X_2}{4} - \frac{X_3}{3} - \frac{X_4}{4}$

III. $y = -1 + X_1 - X_2 + X_3 - X_4$

IV. $y = 0.1X_1 + 0.2X_2 + 0.3X_3 + 0.4X_4$

(A) (I) and (III)

(B) (I) and (IV)

(C) (II) and (IV)

(D) (II) and (III)

Answer: (C)

Solution: The intercept term becomes zero when the observations on study and explanatory variables are standardized in any multiple linear regression model. Since the models in (II) and (IV) do not have intercept term, so they represent the two possible fitted models with standardized observations.