Multiple Linear Regression

Part I: Practice Questions

You do not need to submit these questions.

1. Setup the $\mathbf{X}$ matrix and β vector for each of the following regression models. Assume $i = 1, \ldots, 4$.

   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1} X_{i2} + \varepsilon_i$

   Design Matrix:

$$\begin{pmatrix} 1 & x_{11} & x_{11}x_{12} \\ 1 & x_{21} & x_{21}x_{22} \\ 1 & x_{31} & x_{31}x_{32} \\ 1 & x_{41} & x_{41}x_{42} \end{pmatrix}$$

   Beta vector:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

   (b) $\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

   Design Matrix:

$$\begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{pmatrix}$$

   Beta vector:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

2. Compute the trace of the hat matrix.

$$tr(\mathbf{H}) = tr\left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right) = tr\left(\mathbf{X}\mathbf{X}^T(\mathbf{X}^T\mathbf{X})^{-1}\right) = tr(\mathbf{I}_p) = p$$

3. Consider the following regression model

$$\mathbf{Y} = \mathbf{1}_n \beta + \epsilon$$

   Compute the least squares estimator of β and compute the corresponding hat matrix $\mathbf{H}$.

We will obtain the LS estimators by minimizing $||\mathbf{Y} - \mathbf{1}_n\beta||^2$. Taking the derivative with respect to β and setting equal to 0, we have that

$$-2\mathbf{1}_n^T(\mathbf{Y} - \mathbf{1}_n\beta) = 0$$
$$\mathbf{1}_n^T\mathbf{Y} = \mathbf{1}_n^T\mathbf{1}_n\beta$$
$$\hat{\beta} = (\mathbf{1}_n^T\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{Y}$$

Observe here that

$$(\mathbf{1}_n^T\mathbf{1}_n) = [1\ 1\ \ldots\ 1]\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = n \Rightarrow (\mathbf{1}_n^T\mathbf{1}_n)^{-1} = \frac{1}{n}$$

Also,

$$\mathbf{1}_n^T\mathbf{Y} = [1\ 1\ \ldots\ 1]\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{i=1}^n y_i$$

which leads to

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^n Y_i = \bar{Y}$$

4. For a general linear regression model in which $\mathbf{X}$ may or may not have full rank, show that

$$\sum_{i=1}^n \hat{Y}_i(Y_i - \hat{Y}_i) = 0$$

Let $\mathbf{Y}$ be the column vector of the responses and $mathbfr$ the column vector of the residuals. Then,

$$\sum_{i=1}^n \hat{Y}_i(Y_i - \hat{Y}_i) = \sum_{i=1}^n \hat{Y}_i r_i = \hat{\mathbf{Y}}^T\mathbf{r}$$
$$= (\mathbf{HY})^T(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{Y}^T\mathbf{H}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}$$
$$= \mathbf{Y}^T\mathbf{H}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{H}^T\mathbf{HY} = 0,$$

since $\mathbf{H}$ is symmetric and idempotent.

5. Consider the multiple regression model

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

where ϵ_i are uncorrelated, with $\mathbb{E}(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. Assuming that ε_i are independent normal random variables, state the likelihood function and obtain the maximum likelihood estimators of β_1, and β_2.

This is regression through the origin for the MLR case. So, we have:

$$RSS = \sum_{i=1}^n (Y_i - \beta_1 X_{i1} - \beta_2 X_{i2})^2$$

We take derivatives with respect to β_1 and β_2, and set equal to zero. So,

$$\begin{cases} \frac{\partial RSS}{\partial \beta_1} = -2\sum_{i=1}^n \left(Y_i - \beta_1 X_{i1} - \beta_2 X_{i2}\right) X_{i1} = 0 \\ \frac{\partial RSS}{\partial \beta_2} = -2\sum_{i=1}^n \left(Y_i - \beta_1 X_{i1} - \beta_2 X_{i2}\right) X_{i2} = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^n X_{i1} Y_i = \beta_1 \sum_{i=1}^n X_{i1}^2 + \beta_2 \sum_{i=1}^n X_{i1} X_{i2} \\ \sum_{i=1}^n X_{i2} Y_i = \beta_1 \sum_{i=1}^n X_{i1} X_{i2} + \beta_2 \sum_{i=1}^n X_{i2}^2 \end{cases}$$

If we write the system in matrix format, we get

$$\begin{bmatrix} \sum_{i=1}^n X_{i1} Y_i \\ \sum_{i=1}^n X_{i2} Y_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1} X_{i2} \\ \sum_{i=1}^n X_{i1} X_{i2} & \sum_{i=1}^n X_{i2}^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Solving with respect to β_1, β_2, we obtain

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1} X_{i2} \\ \sum_{i=1}^n X_{i1} X_{i2} & \sum_{i=1}^n X_{i2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n X_{i1} Y_i \\ \sum_{i=1}^n X_{i2} Y_i \end{bmatrix}$$

Substituting the inverse of a 2×2 matrix, we get

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \frac{1}{\sum_{i=1}^n X_{i1}^2 \sum_{i=1}^n X_{i2}^2 - \left(\sum_{i=1}^n X_{i1} X_{i2}\right)^2} \begin{bmatrix} \sum_{i=1}^n X_{i2}^2 & -\sum_{i=1}^n X_{i1} X_{i2} \\ -\sum_{i=1}^n X_{i1} X_{i2} & \sum_{i=1}^n X_{i1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n X_{i1} Y_i \\ \sum_{i=1}^n X_{i2} Y_i \end{bmatrix}$$

Solving with respect to β_1 and β_2 we obtain the desired estimators:

$$\hat{\beta_1} = \frac{\sum_{i=1}^n X_{i2}^2 \sum_{i=1}^n X_{i1} Y_i - \sum_{i=1}^n X_{i2} X_{i1} \sum_{i=1}^n X_{i2} Y_i}{\sum_{i=1}^n X_{i1}^2 \sum_{i=1}^n X_{i2}^2 - \left(\sum_{i=1}^n X_{i1} X_{i2}\right)^2}$$

$$\hat{\beta_2} = \frac{\sum_{i=1}^n X_{i1}^2 \sum_{i=1}^n X_{i1} Y_i - \sum_{i=1}^n X_{i2} X_{i1} \sum_{i=1}^n X_{i1} Y_i}{\sum_{i=1}^n X_{i1} X_{i2}^2 \sum_{i=1}^n X_{i2}^2 - \left(\sum_{i=1}^n X_{i1} X_{i2}\right)^2}$$

Assuming that ε_i are independent normal random variables, state the likelihood function and obtain the maximum likelihood estimators of β_1, and β_2. Are these the same as the least squares estimators?

Assuming that the errors are normally distributed, the likelihood function can be written as

$$L(\beta_1, \beta_2, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n \varepsilon_i^2}{2\sigma^2}\right)$$

The log-likelihood, thus, becomes

$$\ell(\beta_1, \beta_2, \sigma^2) = \log L(\beta_1, \beta_2, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \beta_1 X_{i1} - \beta_2 X_{i2})^2$$

Taking derivatives with respect to β_1, β_2, σ^2 yields:

$$\frac{\partial \ell}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum_{i=1}^n \left(Y_i - \beta_1 X_{i1} - \beta_2 X_{i2}\right) X_{i1}$$

$$\frac{\partial \ell}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum_{i=1}^n \left(Y_i - \beta_1 X_{i1} - \beta_2 X_{i2}\right) X_{i2}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{\sum_{i=1}^n \left(Y_i - \beta_1 X_{i1} - \beta_2 X_{i2}\right) X_{i1}}{\sigma^4}$$

Solving the first two equations, we obtain the estimators for β_1 β_2, which are the same as the LS estimators. Substituting these in the third equation, we obtain an estimator for the variance σ^2.

Part II: Homework Questions – to be submitted

1. Setup the **X** matrix and β vector for each of the following regression models. Assume $i = 1, \ldots, 4$.

   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$

   Design Matrix:

$$\begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 \\ 1 & x_{21} & x_{22} & x_{12}^2 \\ 1 & x_{31} & x_{32} & x_{13}^2 \\ 1 & x_{41} & x_{42} & x_{14}^2 \end{pmatrix}$$

   Beta's vector:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

   (b) $\sqrt{Y_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \varepsilon_i$

   Design Matrix:

$$\begin{pmatrix} 1 & x_{11} & \log_{10} x_{12} \\ 1 & x_{21} & \log_{10} x_{22} \\ 1 & x_{31} & \log_{10} x_{32} \\ 1 & x_{41} & \log_{10} x_{42} \end{pmatrix}$$

   Beta's vector:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

2. Consider the Simple Linear Model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

   where $i = 1, \ldots, n$, and $\epsilon_i \sim IID(0, \sigma^2)$. Suppose that the value of the predictor x_i is replaced by $cx_i + d$, where c, d are some non-zero constant. **Show** how are $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}^2$, R^2 and the t-test of $H_0 : \beta_1 = 0$ affected by this change. *Justify your answer.*

   We will denote the transformed variable with a subscript or superscript c, e.g. $X_c = cX + d$.

   (1) Slope Coefficient

$$\hat{\beta}_1^c = \frac{S_{X_c Y}}{S_{X_c X_c}}$$

   We have

$$S_{X_c X_c} = \sum_i (cx_i + d - \bar{x}_c)^2 = \sum_i (cx_i - \frac{1}{n} \sum_j cx_j)^2 = c^2 S_{XX}$$

$$S_{X_c Y} = \sum_i (cx_i - \bar{x}_c)(y_i - \bar{y}) = c \sum_i (x_i - \bar{x})(y_i - \bar{y}) = cS_{XY}$$

So,

$$\hat{\beta}_1^c = \frac{1}{c}\hat{\beta}_1$$

(2) Intercept

$$\hat{\beta}_0^c = \bar{y} - \hat{\beta}_1^c(c\bar{x} + d) = \bar{y} - \frac{1}{c}\hat{\beta}_1(c\bar{x} + d) = \bar{y} - \hat{\beta}_1\bar{x} - \hat{\beta}_1\frac{d}{c} = \hat{\beta}_0 - \hat{\beta}_1\frac{d}{c}$$

(3) Estimated σ^2

$$\hat{\sigma}_c^2 = \frac{1}{n-2}\sum_i r_{i,c}^2 = \frac{1}{n-2}\sum_i (y_i - \hat{y}_i^c)^2$$

Observe that

$$\hat{y}_i^c = \hat{\beta}_0^c + \hat{\beta}_1^c x_i^c = \hat{\beta}_0 - \frac{d}{c}\hat{\beta}_1 + \frac{1}{c}\hat{\beta}_1(cx_i + d) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

which means that

$$\hat{\sigma}_c^2 = \hat{\sigma}^2$$

(4) Coefficient of Determination R^2

Recall that

$$R_c^2 = r_{X_cY}^2 = r_{XY}^2 = R^2$$

(5) t Test Statistic

The t test statistic is

$$t_c = \frac{\hat{\beta}_1^c}{se(\hat{\beta}_1^c)} = \frac{\frac{1}{c}\hat{\beta}_1}{\frac{1}{c}se(\hat{\beta}_1)} = t$$

3. Obtain the maximum likelihood estimators in a simple linear regression model with normal error terms.

$$\begin{aligned}
\ell(\beta_0, \beta_1, \sigma^2) &= \log L(\beta_0, \beta_1, \sigma^)\\
&= \log \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}}\\
&= -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2
\end{aligned}$$

To maximize the log-likelihood, we take the derivatives with respect to β_0, β_1 and σ^2 and we set them equal to 0. This yields the following estimators:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i))^2$$

4. Show that $Cov(\mathbf{r}) = \sigma^2(\mathbf{I}_n - \mathbf{H})$.

$$Cov(\mathbf{r}) = Cov(\mathbf{y} - \hat{\mathbf{y}}) = Cov(\mathbf{y} - \mathbf{H}\mathbf{y}) = (\mathbf{I} - \mathbf{H})Cov(y) = (\mathbf{I} - \mathbf{H})\sigma^2$$

5. Show that if $\mathbf{X}$ has full rank,

$$(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)^T \mathbf{X}^T \mathbf{X}(\hat{\beta} - \beta)$$

holds, and hence deduce that the left side is minimized uniquely when $\beta = \hat{\beta}$.

$$\begin{aligned}
&(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) = (\mathbf{Y} - \mathbf{X}\beta + \mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\beta + \mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}) \\
&= ((\mathbf{Y} - \mathbf{X}\hat{\beta}) - \mathbf{X}(\beta - \hat{\beta}))^T((\mathbf{Y} - \mathbf{X}\hat{\beta}) - \mathbf{X}(\beta - \hat{\beta})) \\
&= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) - (\mathbf{X}(\beta - \hat{\beta}))^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
&\quad - (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{X}(\beta - \hat{\beta}) + (\mathbf{X}(\beta - \hat{\beta}))^T(\mathbf{X}(\beta - \hat{\beta}) \\
&= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)^T\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta)
\end{aligned}$$

The cross product terms vanish due to orthogonality.