Jared Yu

Module 4 Discussion

In a multiple linear regression model, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where x_1 and x_2 are non-random independent variables, and ε is the random error. Now a set of n items give data, $(y_1, x_{11}, x_{21}), \cdots, (y_n, x_{1n}, x_{2n})$, which follow this model. Decompose SS_T into SS_R and SS_{Res} using the "hat" matrix, discuss and **state the assumptions** in your discussion.

Ans:

Using the method of least squares, suppose that $n > (k = 2)$ and it is assumed that the error term ε has $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$, and that the errors are independent. Furthermore, it has been stated in the question that the x terms are fixed values. The multiple regression model can be expressed using matrix notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{2}$$

Using least-squares, the objective is to find the set of $\hat{\boldsymbol{\beta}}$ that minimizes the following equation,

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{3}$$

This can be rewritten as,

$$S(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \tag{4}$$

$$= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \tag{5}$$

Then, to find the least-squares normal equations, the following must be done,

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \tag{6}$$

$$\rightarrow \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{7}$$

Then to isolate the $\hat{\boldsymbol{\beta}}$ term, multiply both sides of equation (7) by $(\mathbf{X}'\mathbf{X})^{-1}$. An important assumption however is that this inverse matrix, $(\mathbf{X}'\mathbf{X})^{-1}$, exists. This will always be possible if the regressors are linearly independent. The result is the following term,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{8}$$

Now, mirroring the structure of equation (1), the formula for $\hat{\mathbf{y}}$ can be expressed as follows,

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}, \tag{9}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, is known as the "hat matrix."

The following are the formulas (based on the Textbook) for the total sum of squares (SS_T), regression sum of squares (SS_R), and residual sum of squares (SS_{Res}),

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n},$$

$$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n},$$

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y},$$

where it must be shown that $SS_T = SS_R + SS_{Res}$. There are alternative formulas for each of these sums of squares which will be shown below.

SS_T:
It will be shown that an equivalent formula for SS_T is $(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})$, where $\bar{\mathbf{y}}$ is a $n \times 1$ vector consisting all of identical $\bar{y} = \sum_{i=1}^{n} y_i$ terms.

$$\begin{aligned}(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}}) &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\bar{\mathbf{y}} - \bar{\mathbf{y}}'\mathbf{y} + \bar{\mathbf{y}}'\bar{\mathbf{y}} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\bar{\mathbf{y}} + \bar{\mathbf{y}}'\bar{\mathbf{y}}\end{aligned}$$

The previous step holds since $\mathbf{y}'\bar{\mathbf{y}}$ and $\bar{\mathbf{y}}'\mathbf{y}$ are both scalars with the same resulting values. So, the transpose of each other is still the same scalar.

Let's first look at $\bar{\mathbf{y}}'\bar{\mathbf{y}}$. It is the inner product of the same vector which is filled with identical $\bar{y}$ terms. Therefore, the result is $\bar{\mathbf{y}}'\bar{\mathbf{y}} = \sum_{i=1}^{n} \bar{y}^2 = n\bar{y}^2$.

Next, let's look at $-2\mathbf{y}'\bar{\mathbf{y}}$. By first ignoring the constant term -2, it can be seen that $\mathbf{y}'\bar{\mathbf{y}} = \sum_{i=1}^{n} y_i\bar{y} = \bar{y}\sum_{i=1}^{n} y_i = \bar{y}n\bar{y} = n\bar{y}^2$. Therefore, $-2\mathbf{y}'\bar{\mathbf{y}} = -2n\bar{y}^2$.

From this it follows that,

$$(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}}) = \mathbf{y}'\mathbf{y} - 2n\bar{y}^2 + n\bar{y}^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2 = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = SS_T.$$

SS_R:
It will be shown that an equivalent formula for SS_R is $(\hat{\mathbf{y}}-\bar{\mathbf{y}})'(\hat{\mathbf{y}}-\bar{\mathbf{y}})$.

$$(\hat{\mathbf{y}}-\bar{\mathbf{y}})'(\hat{\mathbf{y}}-\bar{\mathbf{y}}) = (\hat{\mathbf{y}}'-\bar{\mathbf{y}}')(\hat{\mathbf{y}}-\bar{\mathbf{y}}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} - \hat{\mathbf{y}}'\bar{\mathbf{y}} - \bar{\mathbf{y}}'\hat{\mathbf{y}} + \bar{\mathbf{y}}'\bar{\mathbf{y}}$$

Starting with $\hat{\mathbf{y}}'\hat{\mathbf{y}}$:

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

Next $\bar{\mathbf{y}}'\bar{\mathbf{y}} = \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$ has already been shown in SS_T.

It can be said that $\hat{\mathbf{y}}'\bar{\mathbf{y}}$ and $\bar{\mathbf{y}}'\hat{\mathbf{y}}$ are both scalars where the same transpose logic can be applied as with $\mathbf{y}'\bar{\mathbf{y}}$ and $\bar{\mathbf{y}}'\mathbf{y}$. So, it will be said first that $\hat{\mathbf{y}}'\bar{\mathbf{y}} = \bar{\mathbf{y}}'\hat{\mathbf{y}}$.

$$\hat{\mathbf{y}}'\bar{\mathbf{y}} = \left(\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\bar{\mathbf{y}} = (\mathbf{H}\mathbf{y})'\left(\frac{1}{n}\mathbf{J}\mathbf{y}\right)$$

Note that $\mathbf{J}$ is an $n \times n$ matrix consisting entirely of 1's.

$$= \frac{1}{n}\mathbf{y}'\mathbf{H}\mathbf{J}\mathbf{y}$$

Note that the hat matrix, $\mathbf{H}$, is a symmetric matrix.

$$= \frac{1}{n}\mathbf{y}'\mathbf{J}\mathbf{y}$$

Note that $\mathbf{HJ} = \mathbf{J}$, this will first be explained. Let $\mathbf{1}$ be an $n \times 1$ vector consisting of 1's only.

$$\mathbf{HJ} = \mathbf{H}[\mathbf{1} \quad \mathbf{1} \quad \cdots \quad \mathbf{1}] = [\mathbf{H1} \quad \mathbf{H1} \quad \cdots \quad \mathbf{H1}] = [\mathbf{1} \quad \mathbf{1} \quad \cdots \quad \mathbf{1}] = \mathbf{J}$$

What is shown above is that first the matrix $\mathbf{J}$ is expressed as a vector of vectors, $[\mathbf{1} \quad \mathbf{1} \quad \cdots \quad \mathbf{1}]$. The hat matrix is then multiplied to this to get to $[\mathbf{H1} \quad \mathbf{H1} \quad \cdots \quad \mathbf{H1}]$. This is simply another way to show typical matrix multiplication. The point is that the hat matrix is a projection matrix on the column space of $\mathbf{X}$, the design matrix. As the design matrix, it includes in the first column a vector of 1's, or $\mathbf{1}$. $\mathbf{H}$ times $\mathbf{1}$, i.e. project $\mathbf{1}$ onto the column space of $\mathbf{X}$, is equal to **1** since **1** is already in the column space of **X**. **J** is a $n \times n$ matrix filled with all 1's. The result then is that **H** multiplied by **J** is also equal to **J**.

$$\cdots = \frac{1}{n}\left[\sum_{i=1}^{n} y_i \quad \cdots \quad \sum_{i=1}^{n} y_i\right]\mathbf{y} = \frac{1}{n}\sum_{j=1}^{n} y_j\left(\sum_{i=1}^{n} y_i\right) = \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

Therefore,

$$\hat{\mathbf{y}}'\hat{\mathbf{y}} - \hat{\mathbf{y}}'\bar{\mathbf{y}} - \bar{\mathbf{y}}'\hat{\mathbf{y}} + \bar{\mathbf{y}}'\bar{\mathbf{y}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - 2\frac{(\sum_{i=1}^{n} y_i)^2}{n} + \frac{(\sum_{i=1}^{n} y_i)^2}{n} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = SS_R$$

SS_{Res}:
It will be shown that an equivalent formula for SS_{Res} is $(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$.

$$(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = SS_{Res}$$

This is taken directly from the lecture notes for Module 3C.

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = \mathbf{y}'\mathbf{y} - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 = \mathbf{y}'\mathbf{y} - n\bar{\mathbf{y}}^2 = (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})$$

Now, to show that $SS_T = SS_R + SS_{Res}$, it will begin with $SS_T's$ alternate expression.

$$(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{\mathbf{y}})$$
$$= [(\mathbf{y} - \hat{\mathbf{y}})' + (\hat{\mathbf{y}} - \bar{\mathbf{y}})'][(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})]$$
$$= (\mathbf{y} - \hat{\mathbf{y}})'[(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})] + (\hat{\mathbf{y}} - \bar{\mathbf{y}})'[(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})]$$
$$= \underbrace{(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})}_{SS_{Res}} + (\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) + \underbrace{(\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}})}_{SS_R}$$

Above, it has been shown how SS_T can be decomposed into SS_{Res}, SS_R, and some other terms. Looking closely at these other terms, it can be seen that they look quite similar. Using the same logic as before with taking the transpose of scalars, it can be done again here. Both $(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}})$ and $(\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$ result in scalar values, and each are the transpose of each other. So, it will be treated as if $(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) = (\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$.

$$\cdots = SS_{Res} + SS_R + 2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}})$$

In order for the decomposition to work, it must be shown then that $2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) = 0$. For simplicity and convenience, the 2 will be ignored.

$$(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) = (\mathbf{y} - \mathbf{H}\mathbf{y})'(\mathbf{H}\mathbf{y} - \bar{\mathbf{y}}) = (\mathbf{y}' - \mathbf{y}'\mathbf{H})(\mathbf{H}\mathbf{y} - \bar{\mathbf{y}})$$
$$= \mathbf{y}'\mathbf{H}\mathbf{y} - \mathbf{y}'\bar{\mathbf{y}} - \mathbf{y}'\mathbf{H}'\mathbf{H}\mathbf{y} + \mathbf{y}'\mathbf{H}'\bar{\mathbf{y}}$$
$$= -\mathbf{y}'\bar{\mathbf{y}} + \mathbf{y}'\mathbf{H}'\bar{\mathbf{y}}$$

Note, the rule just used is that the hat matrix is symmetric and idempotent, therefore $\mathbf{H}'\mathbf{H} = \mathbf{H}$.

$$= \mathbf{y}'\mathbf{H}'\bar{\mathbf{y}} - \mathbf{y}'\bar{\mathbf{y}} = \mathbf{y}'(\mathbf{H} - \mathbf{I})\bar{\mathbf{y}}$$

Note that **I** is the $n \times n$ identity matrix.

$$= -\mathbf{y}'(\mathbf{I} - \mathbf{H})\bar{\mathbf{y}} = -(\mathbf{y}' - \hat{\mathbf{y}}')\bar{\mathbf{y}} = -(\mathbf{y} - \hat{\mathbf{y}})'\bar{\mathbf{y}}$$
$$= -(\mathbf{y} - \mathbf{H}\mathbf{y})'\left(\frac{1}{n}\mathbf{J}\mathbf{y}\right) = -\frac{1}{n}(\mathbf{y}' - \mathbf{y}'\mathbf{H}')\mathbf{J}\mathbf{y} = -\frac{1}{n}(\mathbf{y}'\mathbf{J}\mathbf{y} - \mathbf{y}'\mathbf{H}'\mathbf{J}\mathbf{y}) = -\frac{1}{n}(\mathbf{y}'\mathbf{J}\mathbf{y} - \mathbf{y}'\mathbf{H}\mathbf{J}\mathbf{y})$$

Note that it has already been shown earlier that $\mathbf{HJ} = \mathbf{J}$.

$$= -\frac{1}{n}(\mathbf{y}'\mathbf{J}\mathbf{y} - \mathbf{y}'\mathbf{J}\mathbf{y}) = 0$$

So, it follows that,

$$\boxed{SS_T = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) = SS_{Res} + SS_R.}$$

For completeness, SS_R and SS_{Res} will also be expressed in terms of the hat matrix.

Starting with SS_R:

$$SS_R = (\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}})$$

$$= \left(\mathbf{H}\mathbf{y} - \frac{1}{n}\mathbf{J}\mathbf{y}\right)'\left(\mathbf{H}\mathbf{y} - \frac{1}{n}\mathbf{J}\mathbf{y}\right) = \left(\mathbf{y}'\mathbf{H} - \frac{1}{n}\mathbf{y}'\mathbf{J}\right)\left(\mathbf{H}\mathbf{y} - \frac{1}{n}\mathbf{J}\mathbf{y}\right)$$

$$= \mathbf{y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{y}$$

$$= \mathbf{y}'\left[\mathbf{H} - \frac{1}{n}\mathbf{H}\mathbf{J} - \frac{1}{n}\mathbf{J}\mathbf{H} + \frac{1}{n^2}\mathbf{J}^2\right]\mathbf{y}$$

It has been shown that $\mathbf{H}\mathbf{J} = \mathbf{J}$, however it will now be shown that $\mathbf{J}\mathbf{H} = \mathbf{J}$ also.

$$\mathbf{J}\mathbf{H} = ((\mathbf{J}\mathbf{H})')' = (\mathbf{H}'\mathbf{J}')' = (\mathbf{H}\mathbf{J})' = (\mathbf{J})' = \mathbf{J}$$

Note that the rule used is that both matrices are symmetric.

A new term is $\mathbf{J}^2$, it will be shown below.

$$\mathbf{J}^2 = \mathbf{J}\mathbf{J} = \mathbf{J}[\mathbf{1} \quad \mathbf{1} \quad \cdots \quad \mathbf{1}] = [\mathbf{J}\mathbf{1} \quad \mathbf{J}\mathbf{1} \quad \cdots \quad \mathbf{J}\mathbf{1}] = n\mathbf{J}$$

In the above, it has been shown that the result is an $n \times n$ matrix consisting entirely of the value n. This is expressed as $n\mathbf{J}$.

$$\cdots = \mathbf{y}'\left[\mathbf{H} - \frac{1}{n}\mathbf{J} - \frac{1}{n}\mathbf{J} + \frac{n}{n^2}\mathbf{J}\right]\mathbf{y} = \mathbf{y}'\left[\mathbf{H} - \frac{1}{n}\mathbf{J}\right]\mathbf{y}$$

Lastly, looking at SS_{Res}:

$$SS_{Res} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{H}\mathbf{y})'(\mathbf{y} - \mathbf{H}\mathbf{y})$$

$$= (\mathbf{y}' - \mathbf{y}'\mathbf{H})(\mathbf{y} - \mathbf{H}\mathbf{y}) = \mathbf{y}'(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$$

In conclusion,

$$\boxed{SS_T = \mathbf{y}'\left[\mathbf{H} - \frac{1}{n}\mathbf{J}\right]\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = SS_R + SS_{Res}} \blacksquare$$