RSS

For a linear regression model $\mathbf{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the aim is to minimize the error i.e. $\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) = S(\boldsymbol{\beta})$.

Theorem 1

The minimum of $(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})$ *is attained at* $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y}$.

Proof:

We have

$$\begin{aligned}(\mathbf{y}-\mathrm{X}\boldsymbol{\beta})'(\mathbf{y}-\mathrm{X}\boldsymbol{\beta}) &= (\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}}+\mathrm{X}\widehat{\boldsymbol{\beta}}-\mathrm{X}\boldsymbol{\beta})'(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}}+\mathrm{X}\widehat{\boldsymbol{\beta}}-\mathrm{X}\boldsymbol{\beta}) \\ &= (\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}})+(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathrm{X}'\mathrm{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}),\end{aligned}$$

since

$$\mathrm{X}'(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}}) = \mathrm{X}'(\mathbf{y}-\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y}) = 0.$$

It follows that

$$(\mathbf{y}-\mathrm{X}\boldsymbol{\beta})'(\mathbf{y}-\mathrm{X}\boldsymbol{\beta}) \geq (\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}}).$$

This completes the proof.

RSS

The residual sum of squares is defined to be $(\mathbf{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})$.

Theorem 2

Let $Rank(X) = r$. Then $\mathbb{E}\{(\mathbf{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})\} = (n - r)\sigma^2$.

Proof:

We have

$$\mathbb{E}\{(\mathbf{y}-\mathrm{X}\boldsymbol{\beta})(\mathbf{y}-\mathrm{X}\boldsymbol{\beta})'\} = \mathrm{D}(\mathbf{y}) = \sigma^2\mathrm{I}.$$

Thus,

$$\begin{aligned}
\mathbb{E}(\mathbf{y}\mathbf{y}') &= \mathbb{E}(\mathbf{y})\boldsymbol{\beta}'\mathrm{X}' + \mathrm{X}\boldsymbol{\beta}\mathbb{E}(\mathbf{y}') - \mathrm{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathrm{X}' + \sigma^2\mathrm{I} \\
&= \mathrm{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathrm{X}' + \sigma^2\mathrm{I}. \qquad (1)
\end{aligned}$$

We shall use the notation

$$\mathrm{P} = \mathrm{I} - \mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'$$

throughout this and the next chapter. Observe that P is a symmetric, idempotent matrix and $\mathrm{PX} = 0$. These properties will be useful.

Proof:

Now

$$\begin{aligned}
\mathbb{E}\{(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathrm{X}\widehat{\boldsymbol{\beta}})\} &= \mathbb{E}\{(\mathbf{y}-\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y})'(\mathbf{y}-\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y})\} \\
&= \mathbb{E}\{\mathbf{y}'\mathrm{P}\mathbf{y}\} = \mathbb{E}\{\text{trace}(\mathbf{y}'\mathrm{P}\mathbf{y})\} \\
&= \mathbb{E}\{\text{trace}(\mathrm{P}\mathbf{y}'\mathbf{y})\} = \text{trace}[\mathrm{P}\mathbb{E}\{\mathbf{y}\mathbf{y}'\}] \\
&= \sigma^2\text{trace}(\mathrm{P}),
\end{aligned}$$

by (1) and the fact that $\mathrm{P}\mathrm{X} = 0$.

Proof:

Finally,

$$\begin{aligned}\text{trace}(\mathrm{P}) &= n - \text{trace}[\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'] \\ &= n - \text{trace}[(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}] \\ &= n - \text{Rank}[(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}],\end{aligned}$$

since $(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}$ is idempotent. However,

$$\text{Rank}[(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}] = \text{Rank}[(\mathrm{X}'\mathrm{X})] = \text{Rank}[\mathrm{X}] = r,$$

and the proof is complete.

One–way ANOVA model

Consider the model

$$y_{ij} = \alpha_i + \epsilon_{ij},\ i = 1, \dots, k,\ j = 1, \dots, n_i,$$

where ϵ_{ij} are independent with mean 0 and variance σ^2.

- Find $\widehat{\boldsymbol{\alpha}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathbf{y}$.
- Find RSS using the formula $RSS = \mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\alpha}}'\mathrm{X}'\mathbf{y}$.

Restricted Estimator

Consider the usual regression model $\mathbb{E}(\mathbf{y}) = \mathrm{X}\boldsymbol{\beta}$, $D(\mathbf{y}) = \sigma^2\mathrm{I}$. Suppose we have a priori linear restrictions $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$ on the parameters. We assume that $\mathcal{R}(\mathrm{L}) \subset \mathcal{R}(\mathrm{X})$ and that the equation $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$ is consistent.

Theorem 3

The minimum of $(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})$ *subject to* $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$ *is attained at* $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'\{\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'\}^{-}(\mathrm{L}\widehat{\boldsymbol{\beta}} - \mathrm{z})$, *where* $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y}$.

Proof

Since $\mathcal{R}(\mathrm{L}) \subset \mathcal{R}(\mathrm{X})$ and since $\mathcal{R}(\mathrm{X}) = \mathcal{R}(\mathrm{X}'\mathrm{X})$, then $\mathrm{L} = \mathrm{W}\mathrm{X}'\mathrm{X}$ for some W. Let $\mathrm{T} = \mathrm{W}\mathrm{X}'$. Now,

$$\begin{aligned} \mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}' &= \mathrm{W}\mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}\mathrm{W}' \\ &= \mathrm{W}\mathrm{X}'\mathrm{X}\mathrm{W}' \\ &= \mathrm{T}\mathrm{T}'. \end{aligned}$$

Proof

Since $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$ is consistent, $\mathrm{L}\mathbf{v} = \mathrm{z}$ for some $\mathbf{v}$. Thus,

$$\begin{aligned}
\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}\mathrm{z} &= \mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}\mathrm{L}\mathbf{v} \\
&= \mathrm{T}\mathrm{T}'(\mathrm{T}\mathrm{T}')^{-}\mathrm{T}\mathrm{X}\mathbf{v} \\
&= \mathrm{T}\mathrm{X}\mathbf{v} \\
&= \mathrm{L}\mathbf{v} = \mathrm{z}. \qquad (1)
\end{aligned}$$

Proof

Similarly,

$$
\begin{aligned}
\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}\mathrm{L}\widehat{\beta} &= \mathrm{T}\mathrm{T}'(\mathrm{T}\mathrm{T}')^{-}\mathrm{W}\mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y} \\
&= \mathrm{T}\mathrm{T}'(\mathrm{T}\mathrm{T}')^{-}\mathrm{W}\mathrm{X}'\mathbf{y} \\
&= \mathrm{T}\mathrm{T}'(\mathrm{T}\mathrm{T}')^{-}\mathrm{T}\mathbf{y} \\
&= \mathrm{T}\mathbf{y}. \qquad (2)
\end{aligned}
$$

Proof

And

$$\begin{aligned} L\widehat{\boldsymbol{\beta}} &= L(X'X)^{-}X'\mathbf{y} \\ &= WX'X(X'X)^{-}X'\mathbf{y} \\ &= WX'\mathbf{y} \\ &= T\mathbf{y}. \end{aligned} \tag{3}$$

Proof

Using (1), (2), and (3) we see that $\mathrm{L}\widetilde{\boldsymbol{\beta}} = \mathrm{z}$, and therefore $\widetilde{\boldsymbol{\beta}}$ satisfies the restriction $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$.

Now, for any satisfying $\mathrm{L}\boldsymbol{\beta} = \mathrm{z}$,

$$\begin{aligned}(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) &= (\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}} + \mathrm{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}))'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}} + \mathrm{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &= (\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}}) \\ &\quad + (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathrm{X}'\mathrm{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \qquad (4)\end{aligned}$$

Proof

since it can be shown that $(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathrm{X}'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}}) = 0$ as follows:

$$\begin{aligned} \mathrm{X}'\mathrm{X}\widetilde{\boldsymbol{\beta}} &= \mathrm{X}'\mathrm{X}\widehat{\boldsymbol{\beta}} - \mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}(\mathrm{L}\widehat{\boldsymbol{\beta}} - \mathsf{z}) \\ &= \mathrm{X}'\mathbf{y} - \mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}(\mathrm{L}\widehat{\boldsymbol{\beta}} - \mathsf{z}), \end{aligned}$$

since $\mathrm{L}' = \mathrm{X}'\mathrm{X}\mathrm{W}'$. Hence

$$\mathrm{X}'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}}) = \mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}(\mathrm{L}\widehat{\boldsymbol{\beta}} - \mathsf{z}).$$

Proof

Next, since $\mathrm{L}\widetilde{\boldsymbol{\beta}} = \mathrm{L}\boldsymbol{\beta} = \mathrm{z}$, it follows that

$$\begin{aligned}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathrm{X}'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}}) &= (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathrm{L}'(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')^{-}(\mathrm{L}\widehat{\boldsymbol{\beta}} - \mathrm{z}) \\ &= 0.\end{aligned}$$

From (4) it is clear that

$$(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) \geq (\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathrm{X}\widetilde{\boldsymbol{\beta}}).$$

This completes the proof.

Result

Theorem 4

Rank$(\mathrm{L}) =$ *Rank*$(\mathrm{T}) =$ *Rank*$(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}')$.

Proof: Since $\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}' = \mathrm{T}\mathrm{T}'$, then Rank$(\mathrm{L}(\mathrm{X}'\mathrm{X})^{-}\mathrm{L}') =$ Rank$(\mathrm{T}\mathrm{T}') =$ Rank(T). Clearly, Rank$(\mathrm{L}) =$ Rank$(\mathrm{T}\mathrm{X}) \leq$ Rank(T). Since Rank$(\mathrm{X}) =$ Rank$(\mathrm{X}'\mathrm{X})$, then $\mathrm{X} = \mathrm{M}\mathrm{X}'\mathrm{X}$ for some M. Thus $\mathrm{T} = \mathrm{W}\mathrm{X}' = \mathrm{W}\mathrm{X}'\mathrm{X}\mathrm{M}'$. Therefore, Rank$(\mathrm{T}) \leq$ Rank(L), and hence Rank$(\mathrm{T}) =$ Rank(L).

Example

Consider the model $\mathbb{E}(y_i) = \theta_i$ for $i = 1, 2, 3, 4$ where y_i are uncorrelated with variance σ^2. Find $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ under the restriction $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 0$.