Regression 8

Expected Mean Squares in ANOVA for Linear Regression*

It is well known that the expectation of MSE is given by σ^2 for linear regression models. However, a theoretical proof for this property was not provided in the previous chapters. Here, we investigate the ANOVA results for linear regression models and provide rigorous proofs. In addition, we also provide the distributions of SSE and SSR quantities.

1 Expectations

We will obtain the expectations of the quantities from the ANOVA results that we have studied in the previous chapters. We recall the ANOVA results that we have studied in Chapter 6.

ANOVA table

Source	SS	df	MS	F
Regression	SSR	$p-1$	$\text{MSR} = \text{SSR}/(p-1)$	$F = \frac{\text{MSR}}{\text{MSE}}$
Error	SSE	$n-p$	$\text{MSE} = \text{SSE}/(n-p)$	
Total	SSTo	$n-1$		

*The material covered here is different from that of the textbook.

As provided in the previous chapters, the sums of squares in the ANOVA table are decomposed as follows

$$\underbrace{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}_{\text{SSTo}}=\underbrace{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}_{\text{SSR}}+\underbrace{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}_{\text{SSE}}$$

that is,

$$\underbrace{\|\mathbf{Y}-\bar{Y}\mathbf{1}\|^2}_{\text{SSTo}}=\underbrace{\|\hat{\mathbf{Y}}-\bar{Y}\mathbf{1}\|^2}_{\text{SSR}}+\underbrace{\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2}_{\text{SSE}}.$$

Then, due to the *fundamental theorem of ANOVA* which was mentioned in Theorem 7.1, we have

$$E\Big(\frac{\text{SSTo}}{n-1}\Big)=E\Big(\frac{\text{SSR}}{p-1}\Big)=E\Big(\frac{\text{SSE}}{n-p}\Big)=\sigma^2\quad\text{under } H_0.$$

Thus, SSTo$/(n-1)$, SSR$/(p-1)$ and SSE$/(n-p)$ are all unbiased for σ^2 under H_0 (trivially null model $Y_i=\beta_0+\epsilon_i$). It should be noted that SSTo$/\sigma^2$ and SSR$/\sigma^2$ are unbiased particularly under H_0 only, but SSE$/\sigma^2$ is unbiased under either H_0 or H_1. We will obtain the expected values of these quantities regardless of H_0 or H_1. For notational convenience, we denote

$$\text{MSR}=\frac{\text{SSR}}{p-1}\qquad\text{and}\qquad\text{MSE}=\frac{\text{SSE}}{n-p},\tag{8.1}$$

which are called mean squares. The expected values of these mean squares are called expected mean squares (EMS). Before obtaining the expectations of MSR and MSE, we need to show the following theorems and lemmas.

Theorem 8.1. *Consider a vector of responses* $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ *with* $\boldsymbol{\epsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$. *Then we have* $E(\mathbf{Y}'\mathbf{Y})=n\sigma^2+\|\mathbf{X}\boldsymbol{\beta}\|^2$.

Proof. Since $\boldsymbol{\epsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$, that is, ϵ_i are *iid* from $N(0,\sigma^2)$, we have

$$E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon})=E\left(\sum_{i=1}^{n}\epsilon_i^2\right)=\sum_{i=1}^{n}E(\epsilon_i^2)=\sum_{i=1}^{n}\text{Var}(\epsilon_i^2)=n\sigma^2.\tag{8.2}$$

Using $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ and $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, we have

$$\begin{aligned}
E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) &= E\left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right] \\
&= E\left[\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right] \\
&= E(\mathbf{Y}'\mathbf{Y}) - E(\mathbf{Y}')\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'E(\mathbf{Y}) + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= E(\mathbf{Y}'\mathbf{Y}) - \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= E(\mathbf{Y}'\mathbf{Y}) - \|\mathbf{X}\boldsymbol{\beta}\|^2.
\end{aligned}$$

Thus, it is immediate from (8.2) that we have

$$n\sigma^2 = E(\mathbf{Y}'\mathbf{Y}) - \|\mathbf{X}\boldsymbol{\beta}\|^2,$$

which completes the proof. □

Lemma 8.2. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix and* $\mathbf{B}$ *be* $n \times m$*. The matrices in product form inside a trace can be swapped, that is,* $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$.

Proof. Let $(\mathbf{AB})_{ii}$ be the ith diagonal component of $\mathbf{AB}$. Similarly, $(\mathbf{BA})_{jj}$ is the jth diagonal component of $\mathbf{BA}$. Then we have

$$\mathrm{tr}(\mathbf{AB}) = \sum_{i=1}^{m}(\mathbf{AB})_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} = \sum_{j=1}^{n}(\mathbf{BA})_{jj} = \mathrm{tr}(\mathbf{BA}).$$

□

Theorem 8.3. *We have* $E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = p\sigma^2 + \|\mathbf{X}\boldsymbol{\beta}\|^2$*, where* $\mathbf{H}$ *is a hat matrix given by* $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ *and* $\mathbf{X}$ *is a* $n \times p$ *matrix.*

Proof. First, it is immediate from $\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ and $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ that we have

$$\begin{aligned}
\mathrm{Cov}(\mathbf{Y}) &= E\left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\right] \\
&= E\left[\mathbf{Y}\mathbf{Y}' - \mathbf{Y}\boldsymbol{\beta}'\mathbf{X}' - \mathbf{X}\boldsymbol{\beta}\mathbf{Y}' + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right] \\
&= E(\mathbf{Y}\mathbf{Y}') - E(\mathbf{Y})\boldsymbol{\beta}'\mathbf{X}' - \mathbf{X}\boldsymbol{\beta}E(\mathbf{Y}') + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' \\
&= E(\mathbf{Y}\mathbf{Y}') - \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' = \sigma^2\mathbf{I},
\end{aligned}$$

which results in

$$E(\mathbf{Y}\mathbf{Y}') = \sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'. \tag{8.3}$$

Since $\mathbf{Y}'\mathbf{H}\mathbf{Y}$ is a scalar quantity, we have $\mathbf{Y}'\mathbf{H}\mathbf{Y} = \text{tr}(\mathbf{Y}'\mathbf{H}\mathbf{Y})$. Using Lemma 8.2, we have $\mathbf{Y}'\mathbf{H}\mathbf{Y} = \text{tr}(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = \text{tr}(\mathbf{H}\mathbf{Y}\mathbf{Y}')$. Using this, we have

$$E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = E\left[\text{tr}(\mathbf{H}\mathbf{Y}\mathbf{Y}')\right] = \text{tr}\left[E(\mathbf{H}\mathbf{Y}\mathbf{Y}')\right] = \text{tr}\left[\mathbf{H}E(\mathbf{Y}\mathbf{Y}')\right]. \tag{8.4}$$

Substituting (8.3) into (8.4), we have

$$E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = \text{tr}\left[\mathbf{H}(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')\right] = \text{tr}\left(\sigma^2\mathbf{H} + \mathbf{H}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right).$$

Since $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, we have $\mathbf{H}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' = \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$. Using this, we have

$$E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = \sigma^2\text{tr}(\mathbf{H}) + \text{tr}(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'). \tag{8.5}$$

Using Lemma 8.2, we have

$$\text{tr}(\mathbf{H}) = \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = \text{tr}(\mathbf{I}_p) = p, \tag{8.6}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix due to $\mathbf{X}'\mathbf{X}$ is a $p \times p$ matrix. Similarly, we have

$$\text{tr}(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') = \text{tr}(\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) = \text{tr}[(\mathbf{X}\boldsymbol{\beta})'\mathbf{X}\boldsymbol{\beta}] = \text{tr}(\|\mathbf{X}\boldsymbol{\beta}\|^2) = \|\mathbf{X}\boldsymbol{\beta}\|^2.$$

Substituting these two results into (8.5), we have

$$E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = p\sigma^2 + \|\mathbf{X}\boldsymbol{\beta}\|^2,$$

which completes the proof. □

Theorem 8.4. *We have* $E(\text{SSE}) = (n-p)\sigma^2$.

Proof. In Chapter 5, we have derived

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{H}\mathbf{Y}.$$

It is immediate from Theorems 8.1 and 8.3 that we have

$$E(\text{SSE}) = E(\mathbf{Y}'\mathbf{Y}) - E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) = (n-p)\sigma^2,$$

which completes the proof. □

Theorem 8.4 clearly shows that the MSE in (8.1) is unbiased for σ^2.

Theorem 8.5. *We have*

$$E(\text{SSTo}) = (n-1)\sigma^2 + \left\|(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2$$

and

$$E(\text{SSR}) = (p-1)\sigma^2 + \left\|(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2.$$

Proof. First, we show the result of $E(\text{SSTo})$ whose derivation is very similar to the proof of Theorem 8.3. Since $\mathbf{Y}'\mathbf{J}\mathbf{Y}$ is a scalar quantity, we have $\mathbf{Y}'\mathbf{J}\mathbf{Y} = \text{tr}(\mathbf{Y}'\mathbf{J}\mathbf{Y})$. Using Lemma 8.2, we have $\mathbf{Y}'\mathbf{J}\mathbf{Y} = \text{tr}(\mathbf{Y}'\mathbf{J}\mathbf{Y}) = \text{tr}(\mathbf{J}\mathbf{Y}\mathbf{Y}')$. Using this, we have

$$E(\mathbf{Y}'\mathbf{J}\mathbf{Y}) = E\left[\text{tr}(\mathbf{J}\mathbf{Y}\mathbf{Y}')\right] = \text{tr}\left[E(\mathbf{J}\mathbf{Y}\mathbf{Y}')\right] = \text{tr}\left[\mathbf{J}E(\mathbf{Y}\mathbf{Y}')\right]. \tag{8.7}$$

Substituting $E(\mathbf{Y}\mathbf{Y}') = \sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$ in (8.3) into (8.7), we have

$$\begin{aligned} E(\mathbf{Y}'\mathbf{J}\mathbf{Y}) &= \text{tr}\left[\mathbf{J}(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')\right] = \text{tr}\left(\sigma^2\mathbf{J} + \mathbf{J}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right) \\ &= n\sigma^2 + \text{tr}(\mathbf{J}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') \end{aligned} \tag{8.8}$$

Using Lemma 8.2, we have $\text{tr}(\mathbf{J}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') = \text{tr}(\boldsymbol{\beta}'\mathbf{X}'\mathbf{J}\mathbf{X}\boldsymbol{\beta})$ where $\boldsymbol{\beta}'\mathbf{X}'\mathbf{J}\mathbf{X}\boldsymbol{\beta}$ is a scalar quantity. Thus, we have $\text{tr}(\mathbf{J}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') = \boldsymbol{\beta}'\mathbf{X}'\mathbf{J}\mathbf{X}\boldsymbol{\beta} = (\mathbf{X}\boldsymbol{\beta})'\mathbf{J}\mathbf{X}\boldsymbol{\beta}$. Substituting this result into (8.8), we have

$$E(\mathbf{Y}'\mathbf{J}\mathbf{Y}) = n\sigma^2 + (\mathbf{X}\boldsymbol{\beta})'\mathbf{J}\mathbf{X}\boldsymbol{\beta}. \tag{8.9}$$

In Chapter 5, we have shown that

$$\text{SSTo} = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}.$$

Thus, we have

$$E(\text{SSTo}) = E(\mathbf{Y}'\mathbf{Y}) - \frac{1}{n}E(\mathbf{Y}'\mathbf{J}\mathbf{Y}).$$

Substituting the result in Theorem 8.1 and (8.9) into the above, we have

$$\begin{aligned} E(\text{SSTo}) &= n\sigma^2 + \|\mathbf{X}\boldsymbol{\beta}\|^2 - \frac{1}{n}\left[n\sigma^2 + (\mathbf{X}\boldsymbol{\beta})'\mathbf{J}\mathbf{X}\boldsymbol{\beta}\right] \\ &= (n-1)\sigma^2 + \|\mathbf{X}\boldsymbol{\beta}\|^2 - \frac{1}{n}(\mathbf{X}\boldsymbol{\beta})'\mathbf{J}(\mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

It is seen from $\|\mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{X}\boldsymbol{\beta})'(\mathbf{X}\boldsymbol{\beta})$ that we have

$$E(\text{SSTo}) = (n-1)\sigma^2 + (\mathbf{X}\boldsymbol{\beta})'\mathbf{A}_2(\mathbf{X}\boldsymbol{\beta}).$$

where $\mathbf{A}_2 = \mathbf{I} - \frac{1}{n}\mathbf{J}$ and we recognize that $\mathbf{A}_2$ is symmetric and idempotent (see §12.1 of Graybill, 1983). Thus, we have

$$E(\text{SSTo}) = (n-1)\sigma^2 + (\mathbf{A}_2\mathbf{X}\boldsymbol{\beta})'(\mathbf{A}_2\mathbf{X}\boldsymbol{\beta}) = (n-1)\sigma^2 + \|\mathbf{A}_2\mathbf{X}\boldsymbol{\beta}\|^2,$$

which completes the proof of the first part.

Next, we show the result of $E(\text{SSR})$. In Chapter 5, we have shown that

$$\text{SSR} = \mathbf{Y}'\Big(\mathbf{H} - \tfrac{1}{n}\mathbf{J}\Big)\mathbf{Y}.$$

Thus, we have

$$E(\text{SSR}) = E(\mathbf{Y}'\mathbf{H}\mathbf{Y}) - \frac{1}{n}E(\mathbf{Y}'\mathbf{J}\mathbf{Y}).$$

Substituting the result of Theorem 8.3 and (8.9) into the above, we have

$$E(\text{SSR}) = (p-1)\sigma^2 + \|\mathbf{X}\boldsymbol{\beta}\|^2 - \frac{1}{n}(\mathbf{X}\boldsymbol{\beta})'\mathbf{J}(\mathbf{X}\boldsymbol{\beta})$$

which results in

$$E(\text{SSR}) = (p-1)\sigma^2 + \|\mathbf{A}_2\mathbf{X}\boldsymbol{\beta}\|^2.$$

This completes the proof of the second part. $\square$

Using Theorem 8.4, we have shown that the MSE in (8.1) is unbiased for σ^2. However, from Theorem 8.5, we have $E(\text{MSR}) = \sigma^2 + \frac{1}{p-1}\left\|(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2$ which shows that MSR is not unbiased for σ^2 unless a regression model is trivially $Y_i = \beta_0 + \epsilon_i$.

Theorem 8.6. *Suppose that $Y_i = \beta_0 + \sum_{j=1}^{p-1}\beta_j X_{ij} + \epsilon_i$ for $i = 1, 2, \ldots, n$. Then we have*

$$\left\|(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 = \sum_{j=1}^{p-1}\beta_j^2 S_{jj} + 2\sum_{j<k}^{p-1}\beta_j\beta_k S_{jk},$$

where $S_{jk} = \sum_{i=1}^{n}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$ and $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n}X_{ij}$.

Proof. In Chapter 6, we used the following notations

$$\underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix} \text{ and } \underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}.$$

Then we have

$$\frac{1}{n}\mathbf{J}\mathbf{X} = \begin{bmatrix} 1 & \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_{p-1} \\ 1 & \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_{p-1}. \end{bmatrix}$$

Using the above, we have

$$(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X} = \mathbf{X} - \tfrac{1}{n}\mathbf{J}\mathbf{X} = \begin{bmatrix} 0 & X_{11} - \bar{X}_1 & X_{12} - \bar{X}_2 & \cdots & X_{1,p-1} - \bar{X}_{p-1} \\ 0 & X_{21} - \bar{X}_1 & X_{22} - \bar{X}_2 & \cdots & X_{2,p-1} - \bar{X}_{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & X_{n1} - \bar{X}_1 & X_{n2} - \bar{X}_2 & \cdots & X_{n,p-1} - \bar{X}_{p-1} \end{bmatrix}$$

and

$$(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \sum_{j=1}^{p-1} \beta_j (X_{1j} - \bar{X}_j) \\ \sum_{j=1}^{p-1} \beta_j (X_{2j} - \bar{X}_j) \\ \vdots \\ \sum_{j=1}^{p-1} \beta_j (X_{nj} - \bar{X}_j) \end{bmatrix}.$$

Thus, we have

$$\begin{aligned}
\left\|(\mathbf{I}-\tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 &= \sum_{i=1}^{n}\left[\sum_{j=1}^{p-1}\beta_j(X_{ij}-\bar{X}_j)\right]^2 \\
&= \sum_{i=1}^{n}\left[\sum_{j=1}^{p-1}\beta_j^2(X_{ij}-\bar{X}_j)^2 + 2\sum_{j<k}^{p-1}\beta_j\beta_k(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{p-1}\beta_j^2(X_{ij}-\bar{X}_j)^2 + 2\sum_{i=1}^{n}\sum_{j<k}^{p-1}\beta_j\beta_k(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k) \\
&= \sum_{j=1}^{p-1}\beta_j^2\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2 + 2\sum_{j<k}^{p-1}\beta_j\beta_k\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k) \\
&= \sum_{j=1}^{p-1}\beta_j^2 S_{jj} + 2\sum_{j<k}^{p-1}\beta_j\beta_k S_{jk},
\end{aligned}$$

which completes the proof. □

In summary, we have shown that

$$\begin{aligned}
E(\text{SSE}) &= (n-p)\sigma^2 \\
E(\text{SSTo}) &= (n-1)\sigma^2 + \left\|(\mathbf{I}-\tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 \\
E(\text{SSR}) &= (p-1)\sigma^2 + \left\|(\mathbf{I}-\tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2,
\end{aligned}$$

where

$$\left\|(\mathbf{I}-\tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 = \sum_{j=1}^{p-1}\beta_j^2 S_{jj} + 2\sum_{j<k}^{p-1}\beta_j\beta_k S_{jk}.$$

Example 8.1. For the simple linear regression model given by $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, we obtained $\text{SSR} = \hat{\beta}_1^2 \cdot \sum(X_i - \bar{X})^2 = \hat{\beta}_1^2 \cdot S_{xx}$ in Chapter 2. We also have $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$. Thus, we have $E(\hat{\beta}_1^2) = \text{Var}(\hat{\beta}_1) + \left\{E(\hat{\beta}_1)\right\}^2 = \sigma^2/S_{xx} + \beta_1^2$. Using this, we have

$$E(\text{SSR}) = \sigma^2 + \beta_1^2 S_{xx}.$$

This is also easily obtained by using Theorems 8.5 and 8.6 with $p = 2$. ∥

Example 8.2. Section 6.5 of Kutner et al. (2005) provides

$$E(\text{MSR}) = \sigma^2 + \frac{1}{2}\Big[\beta_1^2\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)^2 + \beta_2^2\sum_{i=1}^{n}(X_{i2}-\bar{X}_2)^2 + 2\beta_1\beta_2\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)\Big].$$

This is also easily obtained by using Theorems 8.5 and 8.6 with $p = 3$. Then we have

$$
\begin{aligned}
E(\text{SSR}) &= 2\sigma^2 + \sum_{j=1}^{2} \beta_j^2 S_{jj} + 2\sum_{j<k}^{2} \beta_j\beta_k S_{jk} \\
&= 2\sigma^2 + \beta_1^2 S_{11} + \beta_2^2 S_{22} + 2\beta_1\beta_2 S_{12}.
\end{aligned}
$$

Thus, using this, we can obtain the $E(\text{MSR})$ above. ‖

2 Distributions

In the previous section, we investigated the expectation of the quantities from the ANOVA table. In Chapter 7, we have studied Theorem 7.1 (fundamental theorem of ANOVA) which provides the distribution of SSE and SSR under H_0. In this section, we will study the distributions of SSR in general (regardless of H_0 or H_1). To derive the distributions of the quantities in the ANOVA table, we need to use the following lemma.

Lemma 8.7. *Let V be a k-dimensional subspace of $\mathbb{R}^n$ and $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$. Let $\boldsymbol{\theta} = E(\mathbf{Y})$, $\hat{\mathbf{Y}}_V = \mathbf{P}_V\mathbf{Y}$ and $\boldsymbol{\theta}_V = \mathbf{P}_V\boldsymbol{\theta}$ where $\mathbf{P}_V$ is an orthogonal projection matrix onto the subspace V. Then we have*

(a) $E(\hat{\mathbf{Y}}_V) = \mathbf{P}_V\boldsymbol{\theta} = \boldsymbol{\theta}_V$.

(b) *If* $\text{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$, *then* $\text{Cov}(\hat{\mathbf{Y}}_V) = \sigma^2\mathbf{P}_V$ *and* $E(\|\hat{\mathbf{Y}}_V\|^2) = k\sigma^2 + \|\boldsymbol{\theta}_V\|^2$.

(c) *If* $\mathbf{Y} \sim N(\boldsymbol{\theta}, \sigma^2\mathbf{I})$, *then* $\hat{\mathbf{Y}}_V \sim N(\boldsymbol{\theta}_V, \sigma^2\mathbf{P}_V)$ *and* $\|\hat{\mathbf{Y}}_V\|^2/\sigma^2 \sim \chi^2_{k,\delta}$ *where* $\delta = \|\boldsymbol{\theta}_V\|^2/\sigma^2$ *is a non-centrality parameter.*

Proof. See Theorem 2.5.3 of Stapleton (2009). □

Theorem 8.8. *Suppose that the model* $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *with* $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. *Then we have*

$$\begin{aligned}
\frac{1}{\sigma^2}\mathrm{SSE} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \sim \chi^2_{n-p} \\
\frac{1}{\sigma^2}\mathrm{SSR} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sim \chi^2_{p-1,\delta} \\
\frac{1}{\sigma^2}\mathrm{SSTo} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2_{n-1,\delta}
\end{aligned}$$

where $\delta = \left\|(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 / \sigma^2$.

Proof. First, let $\mathbf{P}_V = \mathbf{I} - \mathbf{H}$ and $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$. It is immediate from Lemma 8.7 that we have $\|(\mathbf{I} - \mathbf{H})\mathbf{Y}\|^2/\sigma^2 \sim \chi^2_{k,\delta}$. Since

$$\|(\mathbf{I} - \mathbf{H})\mathbf{Y}\|^2 = \mathbf{Y}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathrm{SSE},$$

we have $\mathrm{SSE}/\sigma^2 \sim \chi^2_{k,\delta}$, where $\delta = \|(\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta}\|^2$. The dimension of $\mathbf{P}_V$ is determined by its rank so that we have

$$k = \mathrm{rank}(\mathbf{I} - \mathbf{H}) = \mathrm{tr}(\mathbf{I} - \mathbf{H}) = \mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{H}) = n - \mathrm{tr}(\mathbf{H})$$

due to Corollary 10.2.2 of Harville (2008) which states that for any idempotent matrix $\mathbf{A}$, $\mathrm{rank}(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$. Using $\mathrm{tr}(\mathbf{H}) = p$ from (8.6), we have $k = n - p$. Since

$$\boldsymbol{\theta}_V = (\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0},$$

the non-centrality parameter is $\delta = 0$. This completes the proof of the first part.

Next, let $\mathbf{P}_V = \mathbf{H} - \frac{1}{n}\mathbf{J}$. Similar to the above, we have $k = \mathrm{rank}(\mathbf{H} - \frac{1}{n}\mathbf{J}) = p - 1$. Since

$$\boldsymbol{\theta}_V = (\mathbf{H} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta} = \mathbf{H}\mathbf{X}\boldsymbol{\beta} - \tfrac{1}{n}\mathbf{J}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} - \tfrac{1}{n}\mathbf{J}\mathbf{X}\boldsymbol{\beta} = (\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta},$$

the non-centrality parameter is given by $\delta = \left\|(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{X}\boldsymbol{\beta}\right\|^2 / \sigma^2$. The last part is similarly proved. □

Using the above theorem, we can also obtain the non-central F-distribution

$$F = \frac{\mathrm{SSR}/(p-1)}{\mathrm{SSE}/(n-p)} \sim F_\delta(p-1, n-p).$$

This non-central F-distribution can be used to obtain the power of the hypothesis test under the linear regression model.

References

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Wadsworth, Inc., 2nd edition.

Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill, New York, 5th edition.

Stapleton, J. H. (2009). *Linear Statistical Models*. Wiley, New York, second edition.