Analysis of Variance – Multiple Regression

Model: $Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon \qquad \varepsilon \sim NID\left(0, \sigma^2\right) \qquad p' = p + 1$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{X}^* \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}^* \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \beta_0 \mathbf{1} + \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \qquad E\{\boldsymbol{\varepsilon}\} = \mathbf{0} \quad V\{\boldsymbol{\varepsilon}\} = \sigma^2 \mathbf{I} \qquad E\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta} = \beta_0 \mathbf{1} + \mathbf{X}^* \boldsymbol{\beta}^* \quad V\{\mathbf{Y}\} = \sigma^2 \mathbf{I}$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y} \qquad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y} = \mathbf{P}\mathbf{Y} \qquad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \left(\mathbf{I} - \mathbf{P}\right)\mathbf{Y} \qquad \overline{\mathbf{Y}} = \overline{Y}\mathbf{1} = \left(\frac{1}{n}\right)\mathbf{J}\mathbf{Y}$$

$$V\left\{\hat{\boldsymbol{\beta}}\right\} = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \qquad V\left\{\hat{\mathbf{Y}}\right\} = \sigma^2 \mathbf{P} \qquad V\{\mathbf{e}\} = \sigma^2 \left(\mathbf{I} - \mathbf{P}\right)$$

Total (Corrected) Sum of Squares: $TSS = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = \mathbf{Y}'\left(\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y}$

$$df_{\text{Tot}} = \text{rank}\left(\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right) = tr\left(\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right) = n - 1$$

Error (Residual) Sum of Squares: $SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \mathbf{Y}'\left(\mathbf{I} - \mathbf{P}\right)\mathbf{Y}$

$$df_{\text{Err}} = tr\left(\mathbf{I} - \mathbf{P}\right) = tr\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right) = tr\left(\mathbf{I}\right) - tr\left(\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right) = n - p'$$

Regression Sum of Squares: $SSR = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 = \mathbf{Y}'\left(\mathbf{P} - \left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y} \qquad df_{\text{Reg}} = tr\left(\mathbf{P} - \left(\frac{1}{n}\right)\mathbf{J}\right) = p' - 1 = p$

Distribution of $\frac{SSE}{\sigma^2}$: $\quad \frac{SSE}{\sigma^2} = \mathbf{Y}'\left(\frac{\mathbf{I} - \mathbf{P}}{\sigma^2}\right)\mathbf{Y} = \mathbf{Y}'\mathbf{A}_E\mathbf{Y} \quad \Rightarrow \quad \mathbf{A}_E V\{Y\} \mathbf{A}_E V\{Y\} = \left(\frac{\mathbf{I} - \mathbf{P}}{\sigma^2}\right)\sigma^2\mathbf{I}\left(\frac{\mathbf{I} - \mathbf{P}}{\sigma^2}\right)\sigma^2\mathbf{I} = \left(\frac{\mathbf{I} - \mathbf{P}}{\sigma^2}\right)\sigma^2\mathbf{I}$

$$\Rightarrow \quad \frac{SSE}{\sigma^2} \sim \chi^2 \text{ with } df_{\text{Err}} = \text{rank}\left(\frac{\mathbf{I} - \mathbf{P}}{\sigma^2}\right) = n - p' \text{ and } \Omega_{\text{Err}} = \frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left(\mathbf{I} - \mathbf{P}\right)\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

Distribution of $\frac{SSR}{\sigma^2}$: $\quad \frac{SSR}{\sigma^2} = \mathbf{Y}'\left(\frac{\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}}{\sigma^2}\right)\mathbf{Y} = \mathbf{Y}'\mathbf{A}_R\mathbf{Y}$

$$\Rightarrow \quad \mathbf{A}_R V\{Y\}\mathbf{A}_R V\{Y\} = \left(\frac{\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}}{\sigma^2}\right)\sigma^2\mathbf{I}\left(\frac{\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}}{\sigma^2}\right)\sigma^2\mathbf{I} = \left(\frac{\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}}{\sigma^2}\right)\sigma^2\mathbf{I}$$

$\Rightarrow \quad \frac{SSR}{\sigma^2} \sim \chi^2$ with $df_{\text{Reg}} = \text{rank}\left(\frac{\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}}{\sigma^2}\right) = p$ and

$$\Omega_{\text{Reg}} = \frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left(\mathbf{P}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{X}\boldsymbol{\beta} = \frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{X}\boldsymbol{\beta}$$

Now write: $\frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{X}\boldsymbol{\beta} = \frac{1}{2\sigma^2}\begin{bmatrix}\beta_0 & \boldsymbol{\beta}^{*\prime}\end{bmatrix}\begin{bmatrix}\mathbf{1}' \\ \mathbf{X}^{*\prime}\end{bmatrix}\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\begin{bmatrix}\mathbf{1} & \mathbf{X}^*\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}^*\end{bmatrix} =$

$$\frac{1}{2\sigma^2}\left(\beta_0\mathbf{1}'+\boldsymbol{\beta}^{*\prime}\mathbf{X}^{*\prime}\right)\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\left(\beta_0\mathbf{1}+\mathbf{X}^*\boldsymbol{\beta}^*\right)$$

Note that: $\mathbf{1}'\mathbf{I} = \mathbf{1}'$, $\quad \mathbf{1}'\left(\frac{1}{n}\right)\mathbf{J} = \mathbf{1}'$ and $\mathbf{I}\mathbf{1} = \mathbf{1}$, $\quad \left(\frac{1}{n}\right)\mathbf{J}\mathbf{1} = \mathbf{1} \quad \Rightarrow \quad \mathbf{1}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right) = \mathbf{0}$ and $\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{1} = \mathbf{0}$

$$\Rightarrow \quad \Omega_{\text{Reg}} = \frac{1}{2\sigma^2}\boldsymbol{\beta}^{*\prime}\mathbf{X}^{*\prime}\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{X}^*\boldsymbol{\beta}^*$$

Now, consider for vector $\mathbf{w}$:

$$\mathbf{w}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{w} = \mathbf{w}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{w} = \left(\mathbf{w}-\overline{\mathbf{w}}\right)'\left(\mathbf{w}-\overline{\mathbf{w}}\right)$$

$$= \begin{bmatrix}w_1-\overline{w} & w_2-\overline{w} & \cdots & w_n-\overline{w}\end{bmatrix}\begin{bmatrix}w_1-\overline{w} \\ w_2-\overline{w} \\ \vdots \\ w_n-\overline{w}\end{bmatrix} = \sum_{i=1}^{n}\left(w_i-\overline{w}\right)^2 = 0 \Leftrightarrow \mathbf{w} = \overline{\mathbf{w}}$$

In the case of SSR: $\mathbf{X}^*\boldsymbol{\beta}^* = \begin{bmatrix} \beta_1 X_{11} + \ldots + \beta_p X_{1p} \\ \beta_1 X_{21} + \ldots + \beta_p X_{2p} \\ \vdots \\ \beta_1 X_{n1} + \ldots + \beta_p X_{np} \end{bmatrix}$ $\quad \overline{\mathbf{X}^*\boldsymbol{\beta}^*} = \begin{bmatrix} \beta_1 \overline{X}_1 + \ldots + \beta_p \overline{X}_p \\ \beta_1 \overline{X}_1 + \ldots + \beta_p \overline{X}_p \\ \vdots \\ \beta_1 \overline{X}_1 + \ldots + \beta_p \overline{X}_p \end{bmatrix}$ $\quad \mathbf{X}^*\boldsymbol{\beta}^* - \overline{\mathbf{X}^*\boldsymbol{\beta}^*} = \begin{bmatrix} \sum_{j=1}^{p} \beta_j \left(X_{1j} - \overline{X}_j \right) \\ \sum_{j=1}^{p} \beta_j \left(X_{2j} - \overline{X}_j \right) \\ \vdots \\ \sum_{j=1}^{p} \beta_j \left(X_{nj} - \overline{X}_j \right) \end{bmatrix}$

$$\Rightarrow \quad \Omega_{\text{Reg}} = 0 \quad \Leftrightarrow \quad \sum_{j=1}^{p} \beta_j \left(X_{ij} - \overline{X}_j \right) = 0 \quad i = 1, \ldots, n$$

Case 1: $\beta_1 = \ldots = \beta_p = 0 \quad \Rightarrow \quad \sum_{j=1}^{p} \beta_j \left(X_{ij} - \overline{X}_j \right) = 0 \quad i = 1, \ldots, n$

Case 2: $\beta_k \neq 0, \quad \beta_{k'} = 0 \quad k' \neq k \quad \Rightarrow \quad \beta_k \left(X_{ik} - \overline{X}_k \right) = 0 \quad i = 1, \ldots, n$ which means no variation in X_k

Case 3: $\beta_{k_1}, \ldots, \beta_{k_r} \neq 0 \quad \Rightarrow \quad \sum_{j=1}^{r} \beta_{k_j} \left(X_{ik_j} - \overline{X}_{k_j} \right) = 0 \quad i = 1, \ldots, n$

$$\Rightarrow \quad X_{ik_j} = -\sum_{\substack{j'=1 \\ j' \neq j}}^{r} \frac{\beta_{k_{j'}}}{\beta_{k_j}} \left(X_{ik_{j'}} - \overline{X}_{k_{j'}} \right) + \overline{X}_{k_j} = -\sum_{\substack{j'=1 \\ j' \neq j}}^{r} \frac{\beta_{k_{j'}}}{\beta_{k_j}} X_{ik_{j'}} + \sum_{j'=1}^{r} \frac{\beta_{k_{j'}}}{\beta_{k_j}} \overline{X}_{k_{j'}} \quad i = 1, \ldots, n$$

which implies linear dependencies among the columns of $\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{X}^* \end{bmatrix}$.