Chapter 4 One-Way ANOVA

Hakbae Lee

The Department of Statistics and Data Science, Yonsei University

One-Way ANOVA

- General form of One-Way ANOVA model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i =, \ldots, a, \quad j = 1, \ldots, N_i$$

  where $n = \sum_{i=1}^{a} N_i$, $\mathrm{E}(\epsilon_{ij}) = 0$,$\mathrm{Var}(\epsilon_{ij}) = \sigma^2$ and $\mathrm{Cov}(\epsilon_{ij}, \epsilon_{i'j'}) = 0$ when $(i, j) \neq (i', j')$.

- $\alpha_i = i$-treatment(group) effect

- Balanced model: $N_i = b$ for all i

- Unbalanced model: N_i's are different for all i

More About Models

Example 4.1.1: $a = 3, N_1 = 5, N_2 = 3, N_3 = 3,$

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon = \begin{pmatrix} J_5 & J_5 & 0 & 0 \\ J_3 & 0 & J_3 & 0 \\ J_3 & 0 & 0 & J_3 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{33} \end{pmatrix}$$

Let $N_1 = N_2 = N_3 = 5$. Then

$$\mathbf{X} = (J_3 \otimes J_5, I_3 \otimes J_5)$$

In general, balanced design such as $i = 1, \ldots, a$, $j = 1, \ldots, b$:

$$\mathbf{X} = (J_a \otimes J_b, I_a \otimes J_b)$$

Notation: $\mathbf{J}_r^c \equiv J_r J_c^T = J_r \otimes J^c$ is a $r \times c$ matrix of 1s.

More About Models

Let $\mathbf{Z}$ be the model matrix for the alternative one-way analysis of variance model

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i =, \ldots, a, \quad j = 1, \ldots, N_i$$

Then, letting $X_i^T X_j = \delta_{ij}$ with 1 for $i = j$ and 0 for $i \neq j$,

$$\begin{aligned}
\mathbf{Z} &= (X_1, \ldots, X_a) \\
\mathbf{X} &= [J, \mathbf{Z}] \\
\mathcal{C}(\mathbf{X}) &= \mathcal{C}(\mathbf{Z}) \\
\mathbf{Z}^T\mathbf{Z} &= \text{diag}(N_1, N_2, \ldots, N_a) \\
\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T &= \text{Blk diag}[N_i^{-1}\mathbf{J}_{N_i}^{N_i}] \\
M &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
M_\alpha &= \mathbf{Z}_*(\mathbf{Z}_*^T\mathbf{Z}_*)^{-1}\mathbf{Z}_*^T = M - M_J = M - \frac{1}{n}\mathbf{J}_n^n \\
\mathbf{Z}_* &= (I - M_J)\mathbf{Z} \\
M &= M_J + M_\alpha
\end{aligned}$$

Table 4.1 One-Way Analysis of Variance Table

Matrix Notation		
Source	df	SS
Grand Mean	1	$Y'\left(\frac{1}{n}J_n^n\right)Y$
Treatments	$t-1$	$Y'\left(M-\frac{1}{n}J_n^n\right)Y$
Error	$n-t$	$Y'(I-M)Y$
Total	n	$Y'Y$
Source	SS	$\mathrm{E}(MS)$
Grand Mean	$SSGM$	$\sigma^2+\beta'X'\left(\frac{1}{n}J_n^n\right)X\beta$
Treatments	$SSTrts$	$\sigma^2+\beta'X'\left(M-\frac{1}{n}J_n^n\right)X\beta/(t-1)$
Error	SSE	σ^2
Total	$SSTot$	

Algebraic Notation		
Source	df	SS
Grand Mean	$dfGM$	$n^{-1}y_{\cdot\cdot}^2 = n\bar{y}_{\cdot\cdot}^2$
Treatments	$dfTrts$	$\sum_{i=1}^t N_i\left(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot}\right)^2$
Error	dfE	$\sum_{i=1}^t\sum_{j=1}^{N_i}\left(y_{ij}-\bar{y}_{i\cdot}\right)^2$
Total	$dfTot$	$\sum_{i=1}^t\sum_{j=1}^{N_i}y_{ij}^2$
Source	MS	$\mathrm{E}(MS)$*
Grand Mean	$SSGM$	$\sigma^2+n(\mu+\bar{\alpha}_\cdot)^2$
Treatments	$SSTrts/(t-1)$	$\sigma^2+\sum_{i=1}^t N_i\left(\alpha_i-\bar{\alpha}_\cdot\right)^2/(t-1)$
Error	$SSE/(n-t)$	σ^2
Total		

*$\bar{\alpha}_\cdot = \sum_{i=1}^t N_i\alpha_i/n$

Estimating and Testing Contrasts

A contrast in the one-way ANOVA

$$\lambda^T \beta = \sum_{i=1}^{a} \lambda_i \alpha_i \quad \text{with} \quad \lambda^T J_{a+1} = \sum_{i=1}^{a} \lambda_i = 0$$

For estimable $\lambda^T \beta$, find ρ so that $\rho^T X = \lambda^T$

$$\rho^T = \left(J_{N_i}^T \lambda_i / N_i \right)$$

Proposition 4.2.1. $\lambda^T \alpha = \rho^T \mathbf{X} \beta$ is a contrast if and only if $\rho^T \mathbf{J} = 0$.

Proposition 4.2.2. $\lambda^T \alpha = \rho^T \mathbf{X} \beta$ is a contrast if and only if $\mathrm{M}\rho \in \mathcal{C}(M_\alpha)$.

Since $\sum_{i=1}^{a} \lambda_i = 0$,

$$\sum_{i=1}^{a} \lambda_i \hat{\alpha}_i = \sum_{i=1}^{a} \lambda_i \{\hat{\mu} + \hat{\alpha}_i\} = \sum_{i=1}^{a} \lambda_i \bar{y}_{i+}$$

because $\mu + \alpha_i$ is estimable and its unique LSE is $\bar{y}_{i+}$.

At the significance level α, $H_0 : \lambda^T \alpha = 0$ is rejected if

$$F = \frac{\left(\sum_{i=1}^{a} \lambda_i \bar{y}_{i+}\right)^2 / \left(\sum_{i=1}^{a} \lambda_i^2 / N_i\right)}{MSE} > F(1-\alpha, 1, dfE)$$

Equivalently,

$$t = \frac{\left|\sum_{i=1}^{a} \lambda_i \bar{y}_{i+}\right|}{\sqrt{MSE\left(\sum_{i=1}^{a} \lambda_i^2 / N_i\right)}} > t(1-\frac{\alpha}{2}, dfE)$$

Cochran's Theorem

Let $A_1, \ldots, A_m$ be $n \times n$ symmetric matrices, and $A = \sum_{j=1}^{m} A_j$ with $\texttt{rank}(A_j) = n_j$. Consider the following four statements:

- A_j is an orthogonal projection for all j.
- A is an orthogonal projection (possibly $A = I$).
- $A_j A_k = 0$ for all $j \neq k$.
- $\sum_{j=1}^{m} n_j = n$

If any two of these conditions hold, then all four hold.

Note: Cochran's theorem is a standard result that is the basis of the analysis of variance. If we can write the total sum of squares as a sum of sum of squares components, and if the degrees of freedom add up, then the A_j must be projections, they are orthogonal to each other, and they jointly span R^n