Analysis of Variance Models

1 Non-Full-Rank Models

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $E(\boldsymbol{\epsilon}) = 0$ and $var(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$, but $\boldsymbol{X}$ does not have a full column rank, i.e., $rank(\boldsymbol{X}) = k < p \leq n$ and $\boldsymbol{X}'\boldsymbol{X}$ is singular. In this model, the p parameters in $\boldsymbol{\beta}$ are not unique.

Example 1.1 *One-way Model:*

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{pmatrix}.$$

2 Estimation

2.1 Estimability of $\boldsymbol{\beta}$

The least squares approach lead to solving the normal equations

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

Theorem 2.1 *If $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, the system of equations $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$ is consistent.*

PROOF: The system is consistent if and only if ($\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{c} = \boldsymbol{c}$)

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{y}.$$

Since $\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' = \boldsymbol{X}'$, the system is consistent. $\square$

Since the normal equations are consistent, one solution is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$. Some general results are given below.

(1) $\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{y}$.

(2) $E(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'E(\boldsymbol{y}) = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$, which depends on $(\boldsymbol{X}'\boldsymbol{X})^{-}$ and it is biased.

3. $\boldsymbol{\beta}$ is not estimable. Suppose a linear function $\boldsymbol{A}\boldsymbol{y}$ estimates $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} = E(\boldsymbol{A}\boldsymbol{y}) = E(\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\epsilon}) = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}.$$

Since this must hold for all $\boldsymbol{\beta}$, we must have $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{I}_p$. But $rank(\boldsymbol{A}\boldsymbol{X}) \leq rank(\boldsymbol{X}) < p$. Hence, $\boldsymbol{A}\boldsymbol{X} \neq \boldsymbol{I}_p$, and no such an $\boldsymbol{A}$ exists.

2.2 Estimable function of $\boldsymbol{\beta}$

Definition 2.1 *A linear function of parameters $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is said to be estimable if there exists a linear combination of the observations with an expected value equal to $\boldsymbol{\lambda}'\boldsymbol{\beta}$, i.e., $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if there exists a vector $\boldsymbol{a}$ such that $E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$.*

Theorem 2.2 *In the model* $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, *where* $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ *and* $\boldsymbol{X}$ *is* $n \times p$ *of rank* $k < p \leq n$, *the linear function* $\boldsymbol{\lambda}'\boldsymbol{\beta}$ *is estimable if and only if any one of the following conditions holds:*

(i) $\boldsymbol{\lambda}'$ *is a linear combination of the row of* $\boldsymbol{X}$, *i.e., there exists a vector* $\boldsymbol{a}$ *such that*

$$\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{\lambda}'.$$

(ii) $\boldsymbol{\lambda}'$ *is a linear combination of the rows of* $\boldsymbol{X}'\boldsymbol{X}$ *or* $\boldsymbol{\lambda}$ *is a linear combination of the columns of* $\boldsymbol{X}'\boldsymbol{X}$, *i.e., there exists a vector* $\boldsymbol{r}$ *such that*

$$\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\lambda}' \quad \text{or} \quad \boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}.$$

(iii) $\boldsymbol{\lambda}$ *or* $\boldsymbol{\lambda}'$ *is such that*

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda} \quad \text{or} \quad \lambda'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \lambda',$$

where $(\boldsymbol{X}'\boldsymbol{X})^{-}$ *is any (symmetric) generalized inverse of* $\boldsymbol{X}'\boldsymbol{X}$.

Proof: We only prove the "if" part. (i) If there exists a vector $\boldsymbol{a}$ such that $\boldsymbol{\lambda}' = \boldsymbol{a}'\boldsymbol{X}$, then

$$E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'E(\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{X\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

(ii) If there exists a solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$, then by defining $\boldsymbol{a} = \boldsymbol{X}\boldsymbol{r}$, we have

$$E(\boldsymbol{a}'\boldsymbol{y}) = E(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{r}'\boldsymbol{X}'E(\boldsymbol{y}) = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

(iii) If $\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda}$, then $(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\lambda}$ is a solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \lambda$ in part (ii).

Conversely, if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$ has a solution vector, which can be found as $\boldsymbol{r} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda}$. Substitution into $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$ gives that $\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda}$. $\square$

Thus we can examine linear combinations of the rows of $\boldsymbol{X}$ or of $\boldsymbol{X}'\boldsymbol{X}$ to see what functions of the parameters are estimable.

Example 2.1 *Consider*

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

For example, row(3)-row(1), we have $\boldsymbol{\lambda}' = (0, -1, 1, 0, 0)$, thus $\boldsymbol{\lambda}'\beta = -\alpha_1 + \alpha_2$ is estimable.

We take linear combinations $\mathbf{a}'\boldsymbol{X}$ of the rows of $\boldsymbol{X}$ to obtain three linearly independent rows. Subtracting the first rows from each succeeding rows in $\boldsymbol{X}$, then subtracting the second and third rows from the fourth row of the matrix yields

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus we have three linearly independent estimable functions

$$\boldsymbol{\lambda}_1'\boldsymbol{\beta} = \mu + \alpha_1 + \beta_1, \quad \boldsymbol{\lambda}_2'\boldsymbol{\beta} = \beta_2 - \beta_1, \quad \boldsymbol{\lambda}_3'\boldsymbol{\beta} = \alpha_2 - \alpha_1.$$

3 Estimators

3.1 Estimators of $\lambda'\beta$

From theorem 2.2 (i) and (ii) we have the estimators $\boldsymbol{a}'\boldsymbol{y}$ and $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ for $\boldsymbol{\lambda}'\boldsymbol{\beta}$, where $\boldsymbol{a}'$ and $\boldsymbol{r}'$ satisfy $\boldsymbol{\lambda}' = \boldsymbol{a}'\boldsymbol{X}$ and $\boldsymbol{\lambda}' = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}$, respectively. A third estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a solution of $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$.

Theorem 3.1 *Let $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be an estimable function of $\boldsymbol{\beta}$ in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$. Let $\hat{\boldsymbol{\beta}}$ be any solution to the normal equations $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$, and let $\boldsymbol{r}$ be any solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$. Then the two estimators $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ have the following properties:*

(i) $E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = E(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$.

(ii) $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ *is equal to* $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ *for any* $\hat{\boldsymbol{\beta}}$ *or any* r.

(iii) $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ *and* $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ *are invariant to the choice of* $\hat{\boldsymbol{\beta}}$ *and* r.

PROOF: (i) Since

$$E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

By theorem 2.2 (iii), $\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\lambda}'$, therefore $E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}})$ becomes

$$E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

By theorem 2.2 (ii),

$$E(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{r}'\boldsymbol{X}'E(\boldsymbol{y}) = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

(ii) By theorem 2.2(ii), if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, $\boldsymbol{\lambda}' = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}$ for some $\boldsymbol{r}$. Multiplying the normal equations $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$ by $\boldsymbol{r}'$ gives

$$\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}.$$

Since $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\lambda}'$, we have

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}.$$

(iii) To show that $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ is invariant to the choice of $\boldsymbol{r}$, let $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ be such that $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r}_1 = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{r}_2 = \boldsymbol{\lambda}$. Then

$$\boldsymbol{r}_1'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{r}_1'\boldsymbol{X}'\boldsymbol{y} \quad \text{and} \quad \boldsymbol{r}_2'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{r}_2'\boldsymbol{X}'\boldsymbol{y}.$$

Since $\boldsymbol{r}_1'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{r}_2'\boldsymbol{X}'\boldsymbol{X}$, we have $\boldsymbol{r}_1'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{r}_2'\boldsymbol{X}'\boldsymbol{y}$. It is clear that each is equal to $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$. $\square$

Example 3.1 *Example 11.3.1 (Rencher (2000), p.273-274).*

Theorem 3.2 *Let $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be an estimable function in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$ and cov$(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$. Let $\boldsymbol{r}$ be any solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$ and let $\hat{\boldsymbol{\beta}}$ be any solution to $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$. Then the variance of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ or of $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ has the following properties:*

(i) $var(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}) = \sigma^2\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \sigma^2\boldsymbol{r}'\boldsymbol{\lambda}$.

(ii) $var(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{\lambda}$.

(iii) $var(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}})$ *is unique, that is, invariant to the choice of* $\boldsymbol{r}$ *or of* $(\boldsymbol{X}'\boldsymbol{X})^-$.

PROOF: (i)

$$var(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{r}'\boldsymbol{X}'cov(\boldsymbol{y})\boldsymbol{X}\boldsymbol{y} = \sigma^2\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \sigma^2\boldsymbol{r}'\boldsymbol{\lambda}.$$

(ii)

$$var(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'cov(\hat{\boldsymbol{\beta}})\boldsymbol{\lambda} = \sigma^2\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{\lambda}.$$

Since $\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\lambda}'$, therefore (ii) is proved.

(iii) Since

$$\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{\lambda} = \boldsymbol{r}'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^-(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{r} = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} = \boldsymbol{r}'\boldsymbol{\lambda},$$

it is only need to show that $\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{\lambda}$ is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^-$. Let $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ be two generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. Then

$$\boldsymbol{X}\boldsymbol{G}_1\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{G}_2\boldsymbol{X}'.$$

Multiplying both sides by $\boldsymbol{a}$ such that $\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{\lambda}'$, we obtain

$$\boldsymbol{a}'\boldsymbol{X}\boldsymbol{G}_1\boldsymbol{X}'\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{G}_2\boldsymbol{X}'\boldsymbol{a}$$

or

$$\boldsymbol{\lambda}_1'\boldsymbol{G}_1\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1'\boldsymbol{G}_1\boldsymbol{\lambda}_2.$$

$\square$

Theorem 3.3 *If $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is an estimable function in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \le n$, then the estimators $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{y}$ are BLUE.*

Proof: Let a linear estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be denoted by $\boldsymbol{a}'\boldsymbol{y}$, where without loss of generality $\boldsymbol{a}'\boldsymbol{y} = \boldsymbol{r}'bX'\boldsymbol{y} + \boldsymbol{c}'\boldsymbol{y}$, that is, $\boldsymbol{a}' = \boldsymbol{r}'\boldsymbol{X}' + \boldsymbol{c}'$, where $\boldsymbol{r}'$ is a solution to $\boldsymbol{\lambda}' = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}$. For unbiasedness we must have

$$\boldsymbol{\lambda}'\boldsymbol{\beta} = E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{c}'\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{c}'\boldsymbol{X})\boldsymbol{\beta}.$$

Thus must hold for all $\boldsymbol{\beta}$, and we therefore have

$$\boldsymbol{\lambda}' = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{c}'\boldsymbol{X}.$$

Since $\boldsymbol{\lambda}' = \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}$, it follows that $\boldsymbol{c}'\boldsymbol{X} = 0'$.

$$\begin{aligned} var(\boldsymbol{a}'\boldsymbol{y}) &= \boldsymbol{a}'cov(\boldsymbol{y})\boldsymbol{a} = \sigma^2\boldsymbol{a}'\boldsymbol{a} \\ &= \sigma^2(\boldsymbol{r}'\boldsymbol{X}' + \boldsymbol{c}')(\boldsymbol{r}\boldsymbol{X} + \boldsymbol{c}) \\ &= \sigma^2(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{X}'\boldsymbol{c} + \boldsymbol{c}'\boldsymbol{X}\boldsymbol{r} + \boldsymbol{c}'\boldsymbol{c}) \\ &= \sigma^2(\boldsymbol{r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{r} + \boldsymbol{c}'\boldsymbol{c}). \end{aligned}$$

Therefore, to minimize $var(\boldsymbol{a}'\boldsymbol{y})$, we have $\boldsymbol{c} = 0$ and that $\boldsymbol{r}'\boldsymbol{X}\boldsymbol{y}$ is BLUE. □

3.2 Estimator of σ^2

An estimator of σ^2 is

$$s^2 = \frac{SSE}{n-k}, \tag{1}$$

where $SSE = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\beta}}$ is any solution to the normal equation $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$, and $k = rank(\boldsymbol{X})$.

Theorem 3.4 *For s^2 defined in equation (1) for the non-full-rank model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ with $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, we have the following properties:*

(i) $E(s^2) = \sigma^2$.

(ii) s^2 is invariant to the choice of $\hat{\boldsymbol{\beta}}$ or to the choice of generalized inverse $(\boldsymbol{X}'\boldsymbol{X})^-$.

Proof: (i) Since $SSE = \boldsymbol{y}'(I - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}')\boldsymbol{y}$, we have

$$
\begin{aligned}
E(SSE) &= tr\{[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'](\sigma^2\boldsymbol{I})\} + \boldsymbol{\beta}'\boldsymbol{X}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}']\boldsymbol{X}\boldsymbol{\beta} \\
&= tr\{[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'](\sigma^2\boldsymbol{I})\} \\
&= \sigma^2\{tr(\boldsymbol{I}) - tr[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}']\} \\
&= (n-k)\sigma^2,
\end{aligned}
$$

where $k = rank(\boldsymbol{X}'\boldsymbol{X}) = rank(\boldsymbol{X})$.

(ii)Since $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'$ is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^-$. □

3.3 Normal Model

For the non-full-rank model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, we now assume that

$$\boldsymbol{y} \sim N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}) \quad \text{or} \quad \boldsymbol{\epsilon} \sim N_n(0, \sigma^2\boldsymbol{I}).$$

With the normality assumption we can obtain maximum likelihood estimators.

Theorem 3.5 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators for $\boldsymbol{\beta}$ and σ^2 are*

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y},$$
$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

PROOF: Exercise. □

Theorem 3.6 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and s^2 (corrected for bias) have the following properties:*

(i) $\hat{\boldsymbol{\beta}}$ *is* $N_p((\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-})$.

(ii) $(n-k)s^2/\sigma^2$ *is* $\chi^2(n-k)$.

(iii) $\hat{\boldsymbol{\beta}}$ *and* s^2 *are independent.*

PROOF: Exercise (the proof is the same with the full-rank case). □

4 Testing Hypothesis

It can be shown that unless a hypothesis can be expressed in terms of estimable functions, it cannot be tested (see Searle 1971, pp.193-196). This leads to the following definition:

Definition 4.1 *A hypothesis is said to be* **testable** *if it can be expressed in terms of estimable functions.*

Typically, a testable hypothesis can be written as

$$H_0: \ \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{t},$$

where $\boldsymbol{C} = (\boldsymbol{c}'_{(1)}, \boldsymbol{c}'_{(2)}, \cdots, \boldsymbol{c}'_{(m)})'$, $\boldsymbol{c}_{(i)}\boldsymbol{\beta} = t_i$, $i = 1, 2, \cdots, m$. We assume that

(1) $\boldsymbol{C}$ has a full row rank, $tank(C) = m$.

(2) $\boldsymbol{c}'_{(i)}\boldsymbol{\beta}$ are estimable for all i.

Theorem 4.1 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, if $\boldsymbol{C}$ is $m \times p$ of rank $m \leq k$ such that $\boldsymbol{C}\boldsymbol{\beta}$ is a set of m linearly independent estimable functions, and if $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$, then*

(i) $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$ is nonsingular and invariant to $(\boldsymbol{X}'\boldsymbol{X})^{-}$.

(ii) $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ is $N_m(\boldsymbol{C}\boldsymbol{\beta}, \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')$.

(iii) $SSH/\sigma^2 = (\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}})/\sigma^2$ is $\chi^2(m, \lambda)$, where $\lambda = (\boldsymbol{C}\boldsymbol{\beta})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}(\boldsymbol{C}\boldsymbol{\beta})/2\sigma^2$.

(iv) $SSE/\sigma^2 = \boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']\boldsymbol{y}/\sigma^2$ is $\chi^2(n-k)$.

(v) SSH and SSE are independent.

PROOF: Since

$$\boldsymbol{C}\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{c}_{(1)}\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{c}_{(m)}\boldsymbol{\beta} \end{pmatrix}$$

is a set of m linearly independent estimable functions, then by theorem 2.2 (iii) we have $\boldsymbol{c}_{(i)}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{c}_{(i)}$ for $i = 1, 2, \cdots, m$. Hence,

$$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{C}. \tag{2}$$

Since $rank(AB) \le rank(A)$, we have

$$rank(\boldsymbol{C}) \le rank(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}') \le rank(\boldsymbol{C}).$$

That is, $rank(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}') = m$. Since $rank(\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}')$, we have

$$\begin{aligned} rank(\boldsymbol{C}) &= rank(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}') \\ &= rank(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'] \\ &= rank(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}') \end{aligned} \tag{3}$$

In the last equality, we use the equality $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{C}$. Thus, $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$ is nonsingular. The invariance of $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$ follows from the invariance of $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$.

(ii)

$$E(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \boldsymbol{C}E(\hat{\boldsymbol{\beta}}) = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

By (2), we have $E(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \boldsymbol{C}\boldsymbol{\beta}$.

$$cov(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \boldsymbol{C}cov(\hat{\boldsymbol{\beta}}\boldsymbol{C}' = \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'.$$

By (3), we have $cov(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$.

Due to that $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{y}$, (ii) is proved.

(iii) By part (ii), $cov(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$. Since

$$\sigma^2[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'/\sigma^2 = \boldsymbol{I},$$

the result is followed.

(iv) Exercise. (v) Exercise. □

Theorem 4.2 *Let* $\boldsymbol{y}$ *be* $N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, *where* $\boldsymbol{X}$ *is* $n \times p$ *of rank* $k < p \leq n$, *and let* $\boldsymbol{C}$, $\boldsymbol{C\beta}$, *and* $\hat{\boldsymbol{\beta}}$ *be defined as in theorem 4.1. Then if* $H_0 : \boldsymbol{C\beta} = 0$ *is true, the statistic*

$$F = \frac{SSH/m}{SSE/(n-k)} \\ = \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}})/m}{SSE/(n-k)}$$

is distributed as $F(m, n-k)$.

Proof: Exercise. □

5 Reparameterization

In reparameterization, we transform the non-full-rank model $\boldsymbol{y} = \boldsymbol{X\beta} + \epsilon$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, to the full-rank-model $\boldsymbol{y} = \boldsymbol{Z\gamma} + \epsilon$, where $\boldsymbol{Z}$ is $n \times n \times k$ of rank k and $\boldsymbol{\gamma} = \boldsymbol{U\beta}$ is a set of k independent estimable functions of $\boldsymbol{\beta}$. Thus, $\boldsymbol{Z\gamma} = \boldsymbol{X\beta}$, and we write

$$\boldsymbol{Z\gamma} = \boldsymbol{ZU\beta} = \boldsymbol{X\beta},$$

where $\boldsymbol{X} = \boldsymbol{ZU}$. Since $\boldsymbol{UU}'$ is nonsingular ($rank(\boldsymbol{UU}') = k$), we have

$$\boldsymbol{ZUU}' = \boldsymbol{XU}',$$

and

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{U}'(\boldsymbol{U}\boldsymbol{U}')^{-1}.$$

Now $\boldsymbol{Z}$ is a full-column rank matrix ($rank(Z) \geq rank(\boldsymbol{X}) = k$), and the results for the full-rank model can then be applied here. Hence, we have (the least square estimators)

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}, \\
s^2 &= \frac{1}{n-k}(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\gamma}})'(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\gamma}}) = \frac{SSE}{n-k}.
\end{aligned}$$

Since $\boldsymbol{Z}\boldsymbol{\gamma} = \boldsymbol{X}\boldsymbol{\beta}$,

$$\boldsymbol{Z}\hat{\boldsymbol{\gamma}} = \boldsymbol{X}\hat{\boldsymbol{\beta}},$$

and therefore

$$SSE = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = (\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\gamma}})'(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\gamma}}).$$

Also, for any estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$, we have

$$\boldsymbol{\lambda}'\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{Z}\boldsymbol{\gamma},$$

hence,

$$\widehat{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}} = \boldsymbol{a}'\boldsymbol{Z}\hat{\boldsymbol{\gamma}}.$$

i.e., the estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is invariant to the reparameterization.

Example 5.1 *We illustrate the reparameterization technique for the model* $y_{ij} = \mu + \tau_i + \epsilon_{ij}, i = 1, 2, j = 1, 2$. *In matrix* *form, the model is*

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{pmatrix}.$$

Since $\boldsymbol{X}$ *has rank 2, there exist two linearly independent estimable functions. We can choose these in many ways, one of which is* $\mu + \tau_1$ *and* $\mu + tau_2$. *Thus*

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} = \boldsymbol{U\beta}.$$

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

It is easy to verify that $\boldsymbol{Z\gamma} = \boldsymbol{X\beta}$ *and* $\boldsymbol{ZU} = \boldsymbol{X}$.

6 Side Conditions

The technique of imposing side conditions provides (linear) constraints on a non-full-rank model such that the parameters unique and individually estimable. Another use for side conditions is to impose constraints on the estimates so as to simplify the normal equations. Note that the side conditions must be nonestimable functions of $\boldsymbol{\beta}$.

The matrix $\boldsymbol{X}$ is $n \times p$ of rank $k < p$. Hence the deficiency in the rank of $\boldsymbol{X}$ is $p - k$. In order for all the parameters to be unique, we must define side conditions that make up this deficiency in rank. Accordingly, we define side conditions $\boldsymbol{T}\boldsymbol{\beta} = \underline{0}$, where $\boldsymbol{T}$ is a $(p-k) \times p$ matrix of rank $p - k$ such that $\boldsymbol{T}\boldsymbol{\beta}$ is a set of nonestimable functions.

Theorem 6.1 *If $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\boldsymbol{T}$ is a $(p-k) \times p$ matrix of rank $p - k$ such that $\boldsymbol{T}\boldsymbol{\beta}$ is a set of nonestimable functions, then there us a unique vector $\hat{\boldsymbol{\beta}}$ that satisfies both $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{T}\hat{\boldsymbol{\beta}} = \underline{0}$.*

PROOF: Combine the two equations, we have

$$\begin{pmatrix} \boldsymbol{y} \\ \underline{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon} \\ \underline{0} \end{pmatrix}.$$

Thus, $\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix}' \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix}$ is $p \times p$ of rank p (nonsingular), and we have

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix}' \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix}]^{-1} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix}' \begin{pmatrix} \boldsymbol{y} \\ b0 \end{pmatrix} \\ &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{X}'\boldsymbol{y}. \end{aligned}$$

$\square$

Example 6.1 *Consider the model* $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $i = 1, 2, j = 1, 2$. *It can be shown that the function* $\tau_1 + \tau_2$ *is not estimable. The side condition* $\tau_1 + \tau_2 = 0$ *can be expressed as* $(0, 1, 1)\boldsymbol{\beta} = 0$, *and* $\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{T}'\boldsymbol{T}$ *becomes*

$$\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{T}'\boldsymbol{T} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

Then

$$(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{T}'\boldsymbol{T})^{-1} = \frac{1}{4} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}.$$

With $\boldsymbol{X}'\boldsymbol{y} = (y_{..}, y_{1.}, y_{2.})$, *we obtain*

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix}, \end{aligned}$$

since $y_{1.} + y_{2.} = y_{..}$ *and* $\bar{y}_{i.} = y_{i.}/2$.

7 Full and Reduced Model Test

Suppose we are interested in testing $H_0 : \beta_1 = \cdots = \beta_q$ in the non-full-rank model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta}$ is $p \times 1$ and $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$. If H_0 is testable, we can find a set of linearly independent estimable functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \cdots, \boldsymbol{\lambda}_t'\boldsymbol{\beta}$ such that

H_0 is equivalent to

$$H_0 : \boldsymbol{\gamma}_1 = \begin{pmatrix} \boldsymbol{\lambda}_1'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\lambda}_t'\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It is also possible to find

$$\boldsymbol{\gamma}_2 = \begin{pmatrix} \boldsymbol{\lambda}_{t+1}'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\lambda}_k'\boldsymbol{\beta} \end{pmatrix}$$

such that k functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \cdots, \boldsymbol{\lambda}_k'\boldsymbol{\beta}$ are linearly independent and estimable, where $k = rank(\boldsymbol{X})$. Let

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{pmatrix}.$$

We can now reparameterize from the non-full-rank model to the full rank model

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} = \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \boldsymbol{Z}_2\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon},$$

where $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ is partitioned to conform with the number of elements in $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$.

Since $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ is a full-rank model, the hypothesis $H_0 : \boldsymbol{\gamma}_1 = \underline{0}$ can be tested as in the full-rank model. The test is outlined in table 1. Note that the degrees of freedom, t, for $SS(\boldsymbol{\gamma}_1|\boldsymbol{\gamma}_2)$ is the number of linearly independent estimable functions required to express H_0.

Table 1: ANOVA for testing $H_0 : \boldsymbol{\gamma}_1 = \underset{\sim}{0}$ in reparameterized balanced models

Source of Variation	d.f.	Sum of Squares	F-Statistic
Due to $\boldsymbol{\gamma}_1$ adjusted for $\boldsymbol{\gamma}_2$	t	$SS(\gamma_1 \| \gamma_2) = \hat{\boldsymbol{\gamma}}'\boldsymbol{Z}'\boldsymbol{y} - \hat{\boldsymbol{\gamma}}_2'\boldsymbol{Z}_2'\boldsymbol{y}$	$\frac{SS(\boldsymbol{\gamma}_1 \| \boldsymbol{\gamma}_2)/t}{SSE/(n-k)}$
Error	$n-k$	$SSE = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\gamma}}'\boldsymbol{Z}'\boldsymbol{y}$	
Total	$n-1$	$SST = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2$	

Table 2: ANOVA for testing $H_0 : \boldsymbol{\gamma}_1 = \underset{\sim}{0}$ in reparameterized balanced models

Source of Variation	d.f.	Sum of Squares	F-Statistic
Due to $\boldsymbol{\beta}_1$ adjusted for $\boldsymbol{\beta}_2$	t	$SS(\boldsymbol{\beta}_1 \| \boldsymbol{\beta}_2) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y}$	$\frac{SS(\boldsymbol{\beta}_1 \| \boldsymbol{\beta}_2)/t}{SSE/(n-k)}$
Error	$n-k$	$SSE = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$	
Total	$n-1$	$SST = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2$	

Sine in the reparameterization model, we have $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}$, we have

$$\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = \hat{\boldsymbol{\gamma}}'\boldsymbol{Z}'\boldsymbol{y},$$

where $\hat{\boldsymbol{\beta}}$ is any solution to the normal equation $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$. Similarly, corresponding to $\boldsymbol{y} = \boldsymbol{Z}_2\boldsymbol{\gamma}_2^* + \boldsymbol{\epsilon}^*$, we have a reduced model $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2^* + \boldsymbol{\epsilon}^*$ obtained by setting $\beta_1 = \cdots = \beta_q$. Then

$$\hat{\boldsymbol{\beta}}_2^{*'}\boldsymbol{X}_2'\boldsymbol{y} = \hat{\boldsymbol{\gamma}}_2^{*'}\boldsymbol{Z}_2'\boldsymbol{y}.$$

The test can then be expressed as in Table 2, in which $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$ is obtained from the full model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y}$ is obtained from the model $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, which has been reduced by the hypothesis $H_0 : \beta_1 = \cdots = \beta_q$.

8 One Way analysis of Variance: Balanced Case

8.1 The one-way model

The one-way balanced model can be expressed as follows:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \cdots, k, \quad j = 1, 2, \cdots, n. \qquad (4)$$

If $\alpha_1, \cdots, \alpha_k$ represent the effects of k treatments, each of which is applied to n experimental units, then y_{ij} is the response of the jth observation among the n units that receive the ith treatment. The assumptions for the model are

(1) $E(\epsilon_{ij}) = 0$ for all i, j.

(2) $var(\epsilon_{ij}) = \sigma^2$ for all i, j.

(3) $cov(\epsilon_{ij}, \epsilon_{rs}) = 0$ for all $(i, j) \neq (r, s)$.

We sometimes have the distribution assumption that

(4) ϵ_{ij} is distributed as $N(0, \sigma^2)$.

In this model, we often use μ_i to denote the mean for the ith treatment, i.e., $E(y_{ij}) = \mu_i$, using assumption (1), we have $\mu_i = \mu + \alpha_i$. We can thus write the model in the form

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, \cdots, k, \quad j = 1, 2, \cdots, n.$$

In this form of the model, the hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ is of interest.

8.2 Estimation of parameters

Extending (4) to a general k and n, the one-way model can be written in matrix form as

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{j} & \boldsymbol{j} & \underline{0} & \cdots & \underline{0} \\ \boldsymbol{j} & \underline{0} & \boldsymbol{j} & \cdots & \underline{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{j} & \underline{0} & \underline{0} & \cdots & \boldsymbol{j} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_k \end{pmatrix}$$

or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{j}$ and $\underline{0}$ are each of size $n \times 1$, and $\boldsymbol{y}_i$ and $\boldsymbol{\epsilon}_i$ are defined as

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix}, \quad \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in} \end{pmatrix}.$$

Thus, the normal equation $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$ takes the form

$$\begin{pmatrix} kn & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ n & 0 & 0 & \cdots & n \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_k \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{k.} \end{pmatrix},$$

where $y_{..} = \sum_{ij} y_{ij}$ and $y_{i.} = \sum_j y_{ij}$.

A generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$ is given by

$$(\boldsymbol{X}'\boldsymbol{X})^{-} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1/n & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/n \end{pmatrix}. \tag{5}$$

Then a solution to the normal equation is obtained as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 0 \\ \bar{y}_{1.} \\ \vdots \\ \bar{y}_{k.} \end{pmatrix}. \tag{6}$$

The estimators in (6) are different for different $(\boldsymbol{X}'\boldsymbol{X})^{-}$, but they give the same estimates of estimable functions, since $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ is invariant to the choice of $\hat{\boldsymbol{\beta}}$.

Using $\hat{\boldsymbol{\beta}}$ in (6), we can express SSE in the following form:

$$\begin{aligned} SSE &= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \sum_{i=1}^{k} \bar{y}_{i.}y_{i.} \\ &= \sum_{ij} y_{ij}^2 - \sum_{i} \frac{y_{i.}^2}{n}. \end{aligned}$$

Thus, s^2 $(E(s^2) = \sigma^2)$ is given by

$$s^2 = \frac{1}{k(n-1)}[\sum_{ij} y_{ij}^2 - \sum_{i} \frac{y_{i.}^2}{n}].$$

8.3 Testing the hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

Using the relationship $\mu_i = \mu + \alpha_i$, the hypothesis can be expressed as $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k$, which is testable because it can be written in terms of $k - 1$ linearly independent estimable contrasts, for example, $H_0 : \alpha_1 - \alpha_2 = \alpha_1 - \alpha_3 = \cdots = \alpha_1 - \alpha_k = 0$.

For simplicity, we illustrate the testing procedure with $k = 4$. In this case, $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_2, \alpha_4)'$ and the hypothesis is $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$. Using three linearly independent estimable contrasts, the hypothesis can be written in the form

$$H_0 : \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \\ \alpha_1 - \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which can be expressed as $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \underline{0}$, where

$$\boldsymbol{C} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}. \tag{7}$$

The matrix $\boldsymbol{C}$ in (7) is not unique, for example,

$$\boldsymbol{C}_2 = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Using $\boldsymbol{C}$ in (7) and $(\boldsymbol{X}'\boldsymbol{X})^-$ in (5), we have

$$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}' = \frac{1}{n}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}. \tag{8}$$

To find the inverse of (8), we write it in the form

$$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}' = \frac{1}{n}\{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\} = \frac{1}{n}(\boldsymbol{I}_3 + \boldsymbol{j}_3\boldsymbol{j}_3').$$

By the formula:

$$(\boldsymbol{B} + \boldsymbol{c}\boldsymbol{c}')^{-1} = \boldsymbol{B}^{-1} - \frac{\boldsymbol{B}^{-1}\boldsymbol{c}\boldsymbol{c}'\boldsymbol{B}^{-1}}{1 + \boldsymbol{c}'\boldsymbol{B}^{-1}\boldsymbol{c}},$$

we have

$$[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}']^{-1} = n(\boldsymbol{I}_3 - \frac{\boldsymbol{I}_3^{-1}\boldsymbol{j}_3\boldsymbol{j}_3'\boldsymbol{I}_3^{-1}}{1 + \boldsymbol{j}_3'\boldsymbol{I}_3^{-1}\boldsymbol{j}_3}) = n(\boldsymbol{I}_3 - \frac{1}{4}\boldsymbol{J}_3), \tag{9}$$

where $\boldsymbol{J}_3$ is 3×3.

In addition, we have

$$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \frac{1}{n}\begin{pmatrix} \boldsymbol{j}_n' & -\boldsymbol{j}_n' & \underline{0}' & \underline{0}' \\ \boldsymbol{j}_n' & \underline{0}' & -\boldsymbol{j}_n' & \underline{0}' \\ \boldsymbol{j}_n' & \underline{0}' & \underline{0}' & -\boldsymbol{j}_n' \end{pmatrix} = \frac{1}{n}\boldsymbol{A}, \tag{10}$$

where $\boldsymbol{j}_n'$ and $\underline{0}'$ are $1 \times n$.

Using (9) and (10), we have

$$\begin{aligned}
SSH &= (\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}\boldsymbol{C}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}']^{-1}\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{y}'[\frac{1}{n}\boldsymbol{A}'n(I_3 - \frac{1}{4}\boldsymbol{J}_3)\frac{1}{n}\boldsymbol{A}]\boldsymbol{y} \\
&= \boldsymbol{y}'[\frac{1}{n}\boldsymbol{A}'\boldsymbol{A} - \frac{1}{4n}\boldsymbol{A}'\boldsymbol{J}_3\boldsymbol{A}]\boldsymbol{y} \\
&= \boldsymbol{y}'[\frac{1}{4n}\begin{pmatrix} 3\boldsymbol{J}_n & -\boldsymbol{J}_n & -\boldsymbol{J}_n & -\boldsymbol{J}_n \\ -\boldsymbol{J}_n & 3\boldsymbol{J}_n & -\boldsymbol{J}_n & -\boldsymbol{J}_n \\ -\boldsymbol{J}_n & -\boldsymbol{J}_n & 3\boldsymbol{J}_n & -\boldsymbol{J}_n \\ -\boldsymbol{J}_n & -\boldsymbol{J}_n & -\boldsymbol{J}_n & 3\boldsymbol{J}_n \end{pmatrix}]\boldsymbol{y}' \\
&= \boldsymbol{y}'[\frac{1}{4n}\boldsymbol{B}]\boldsymbol{y}
\end{aligned}$$

Note that

$$\frac{1}{4n}\boldsymbol{B} = \frac{1}{n}\begin{pmatrix} 4\boldsymbol{J}_n & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & 4\boldsymbol{J}_n & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & 4\boldsymbol{J}_n & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & 4\boldsymbol{J}_n \end{pmatrix} - \frac{1}{4n}\begin{pmatrix} \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n \\ \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n \\ \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n \\ \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n & \boldsymbol{J}_n \end{pmatrix}.$$

Hence, we have

$$\begin{aligned}
SSH &= \frac{1}{n}\sum_{i=1}^{4}\boldsymbol{y}_i'\boldsymbol{J}_n\boldsymbol{y}_i - \frac{1}{4n}\boldsymbol{y}'\boldsymbol{J}_{4n}\boldsymbol{y} \\
&= \frac{1}{n}\sum_{i=1}^{4}\boldsymbol{y}_i'\boldsymbol{j}_n\boldsymbol{j}_n'\boldsymbol{y}_i - \frac{1}{4n}\boldsymbol{y}'\boldsymbol{j}_{4n}\boldsymbol{j}_{4n}'\boldsymbol{y} \\
&= \frac{1}{n}\sum_{i=1}^{4}y_{i.}^2 - \frac{1}{4n}y_{..}^2.
\end{aligned}$$

Table 3: Ascorbic Acid (mg/100g) for three packaging methods

	A	B	C
	14.29	20.06	20.04
	19.10	20.64	26.23
	19.09	18.00	22.74
	16.25	19.56	24.04
	15.09	19.47	23.37
	16.61	19.07	25.02
	19.63	18.38	23.27
Totals ($y_{i.}$)	120.06	135.18	164.71
Means ($\bar{y}_{i.}$)	17.15	19.31	23.53

Example 8.1 *(Ascorbic Acid) Three methods (A-C) of packaging frozen foods were compared by Daniel (1974, p.196). The response variable was ascorbic acid. The data are given in the table 8.1.*

To test the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$, we calculate

$$\frac{y_{..}^2}{kn} = \frac{419.95^2}{(3)(7)} = 8398.0001$$

$$\frac{1}{7}\sum_{i=1}^{3} y_{i.}^2 = \frac{1}{7}[120.06^2 + 135.18^2 + 164.71^2] = 8545.3457$$

$$\sum_{i=1}^{3}\sum_{j=1}^{7} y_{ij}^2 = 8600.3127$$

The sums of squares for the treatments, error, and total are

Table 4: Analysis of Variance for the Ascorbic Acid Data

Source	d.f.	Sum of Squares	Mean Square	F
Method	2	147.3456	73.6728	24.1256
Error	18	54.9670	3.9537	
Total	20	202.312		

then

$$SSH = \frac{1}{7}\sum_{i=1}^{3} y_{i.}^2 - \frac{y_{..}^2}{21} = 8545.3457 - 8398.0001 = 147.3456,$$

$$SSE = \sum_{ij} y_{ij}^2 - \frac{1}{7}\sum_{i} y_{i.}^2 = 8600.3127 - 8545.3457 = 54.9670$$

$$SST = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{21} = 8600.3127 - 8398.0001 = 202.3126$$

These sums of squares can be used to obtain an F-test, as shown in table 8.1. The p-value for $F = 24.1256$ is 8.07×10^{-6}. Thus, we reject $H_0 : \mu_1 = \mu_2 = \mu_3$.

8.4 Hypothesis test for a contrast

In exercises, we have shown that for the one-way balanced model, contrasts in α's are estimable, that is $\sum_i c_i\alpha_i$ is estimable if and only if $\sum_i c_i = 0$. Since

$$\sum_{i=1}^{k} c_i\mu_i = \sum_{i=1}^{k} c_i(\mu + \alpha_i) = \mu\sum_{i=1}^{k} c_i + \sum_{i=1}^{k} c_i\alpha_i = \sum_{i=1}^{k} c_i\alpha_i,$$

the contrast $\sum_i c_i\alpha_i$ is equivalent to $\sum_i c_i\mu_i$.

A hypothesis of interest is

$$H_0 : \sum_{i=1}^{k} c_i \alpha_i = 0 \quad \text{or} \quad H_0 : \sum_{i=1}^{k} c_i \mu_i = 0,$$

which represents a comparison of means if $\sum_i c_i = 0$. For example,

$$H_0 : \ 3\mu_1 - \mu_2 - \mu_3 - \mu_4 = 0$$

can be written as

$$H_0 : \mu_1 = \frac{1}{3}(\mu_2 + \mu_3 + \mu_4),$$

which compares μ_1 with the average of μ_2, μ_3 and μ_4.

The hypothesis can be expressed as $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = 0$, where $\boldsymbol{c}' = (0, c_1, \cdots, c_k)$ and $\boldsymbol{\beta} = (\mu, \alpha_1, \cdots, \alpha_k)'$. Assuming $\boldsymbol{y}$ is $N_{kn}(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$, H_0 can be tested using the following F statistic:

$$\begin{aligned} F &= \frac{(\boldsymbol{c}'\hat{\boldsymbol{\beta}})'[\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}]^{-1}\boldsymbol{c}'\hat{\boldsymbol{\beta}}}{SSE/k(n-1)} \\ &= \frac{(\boldsymbol{c}'\hat{\boldsymbol{\beta}})^2}{s^2 \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}} \\ &= \frac{(\sum_{i=1}^{k} c_i \bar{y}_{i.})^2}{s^2 \sum_{i=1}^{k} c_i^2/n}, \end{aligned}$$

where $s^2 = SSE/k(n-1)$ and $(\boldsymbol{X}'\boldsymbol{X})^{-}$ and $\hat{\boldsymbol{\beta}}$ are given by (5) and (6).

9 Two-Way Analysis of Variance: Balanced Case

Suppose we have the additive (no-interaction) model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \cdots, a, \quad j = 1, \cdots, b$$

This model is two-factor design with balanced data. Factor A (for α) has a levels. Factor B (for β) has b levels. Only one observation y_{ij} in each (i, j) cell.

In matrix form, the model can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} = (y_{11}, \cdots, y_{1b}, y_{21}, \cdots, y_{2b}, \cdots, y_{a1}, \cdots, y_{ab})'$,
$\boldsymbol{\beta} = (\mu, \alpha_1, \cdots, \alpha_a, \beta_1, \cdots, \beta_b)$,
$\boldsymbol{\epsilon} = (\epsilon_{11}, \cdots, \epsilon_{1b}, \epsilon_{21}, \cdots, \epsilon_{2b}, \cdots, \epsilon_{a1}, \cdots, \epsilon_{ab})$, and

$$\boldsymbol{X} = \begin{array}{c} \begin{matrix} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_a & \beta_1 & \beta_2 & \cdots & \beta_b \end{matrix} \\ \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \end{array}$$

$$
\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix}
ab & b & b & \cdots & b & a & a & \cdots & a \\
b & b & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\
b & 0 & b & \cdots & 0 & 1 & 1 & \cdots & 1 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
b & 0 & 0 & \cdots & b & 1 & 1 & \cdots & 1 \\
a & 1 & 1 & \cdots & 1 & a & 0 & \cdots & 0 \\
a & 1 & 1 & \cdots & 1 & 0 & a & \cdots & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
a & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & a
\end{pmatrix}
$$

It is not easy to find a generalized inverse $(\boldsymbol{X}'\boldsymbol{X})^{-}$. Instead, we can impose two side conditions $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$ to solve the normal equation: $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$.

$$
\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{pmatrix}
ab\mu + b\sum_i \alpha_i + a\sum_j \beta_j \\
b(\mu + \alpha_1) + \sum_j \beta_j \\
\vdots \\
b(\mu + \alpha_a) + \sum_j \beta_j \\
a(\mu + \beta_1) + \sum_i \alpha_i \\
\vdots \\
a(\mu + \beta_b) + \sum_i \alpha_i
\end{pmatrix} = \begin{pmatrix}
ab\mu \\
b(\mu + \alpha_1) \\
\vdots \\
b(\mu + \alpha_a) \\
a(\mu + \beta_1) \\
\vdots \\
a(\mu + \beta_b)
\end{pmatrix}
$$

$$
\boldsymbol{X}'\boldsymbol{y} = (y_{..}, y_{1.}, \cdots, y_{a.}, y_{.1}, \cdots, y_{.b})'
$$

So the solution are:

$$
\hat{\boldsymbol{\beta}} = (\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_a, \hat{\beta}_1, \cdots, \hat{\beta}_b)',
$$

where

$$\hat{\mu} = y_{..}/(ab) = \bar{y}_{..},$$
$$\hat{\alpha}_i = y_{i.}/b - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..}, \quad i = 1, \cdots, a$$
$$\hat{\beta}_j = y_{.j}/a - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..}, \quad , j = 1, \cdots, b.$$

We now proceed to obtain the test for $H_0 : \alpha_1 = \cdots = \alpha_a$ following the outline in table 2. The hypothesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ can be expressed as $H_0 : \alpha_1 - \alpha_2 = 0$ and $\alpha_1 - \alpha_3 = 0$. Thus H_0 is testable if $\alpha_1 - \alpha_2$ and $\alpha_1 - \alpha_3$ are estimable. Since each expectation of observation $E(y_{ij}) = \mu + \alpha_i + \beta_j$ is estimable, and any linear combination of $(\mu + \alpha_i + \beta_j)$'s is estimable, $\alpha_1 - \alpha_2$ and $\alpha_1 - \alpha_3$ are both estimable. $(\alpha_1 - \alpha_2 = (\mu + \alpha_1 + \beta_1) - (\mu + \alpha_2 + \beta_1))$

First, we calculate

$$SS(\mu, \alpha, \beta) = \hat{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{y} = (\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_a, \hat{\beta}_1, \cdots, \hat{\beta}_b)' \begin{pmatrix} y_{..} \\ y_{1.} \\ \cdots \\ y_{a.} \\ y_{.1} \\ \cdots \\ y_{.b} \end{pmatrix}$$

$$= \bar{y}_{..}y_{..} + \sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})y_{i.} + \sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})y_{.j}$$

$$= \frac{y_{..}^2}{ab} + (\sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}) + (\sum_{j=1}^{b}\frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}),$$

since $\sum_i y_{i.} = y_{..}$ and $\sum_j y_{.j} = y_{..}$.

The error sum of squares SSE is given by

$$\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{ab} - (\sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}) - (\sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}).$$

To obtain $\hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y}$ in table 2, we use the reduced model $y_{ij} = \mu + \alpha + \beta_j + \epsilon_{ij} = \mu + \beta_j + \epsilon_{ij}$, where $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$ and $\mu + \alpha$ is replaced by μ. The normal equations $\boldsymbol{X}_2'\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{y}$ for the reduced model are

$$\begin{aligned} ab\hat{\mu} + a\hat{\beta}_1 + a\hat{\beta}_2 &= y_{..} \\ a\hat{\mu} + a\hat{\beta}_1 &= y_{.1} \\ a\hat{\mu} + a\hat{\beta}_2 &= y_{.2}. \end{aligned}$$

Using the side condition $\hat{\beta}_1 + \hat{\beta}_2 = 0$, the solution to the reduced normal equations is easily obtained as

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\beta}_1 = \bar{y}_{.1} - \bar{y}_{..}, \quad \cdots, \quad \hat{\beta}_b = \bar{y}_{.b} - \bar{y}_{..}.$$

Thus, we have

$$SS(\mu, \boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y} = \hat{\mu}y_{..} + \hat{\beta}_1 y_{.1} + \cdots + \hat{\beta}_b y_{.b} = \frac{y_{..}^2}{ab} + (\sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}).$$

$$SS(\boldsymbol{\alpha}|\mu, \boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y} = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}.$$

The test is summarized in table 5. The test statistic for $H_0 : \beta_1 = \cdots = \beta_b$ can be obtained similarly.

Table 5: ANOVA for two-way models

Source of Variation	d.f.	Sum of Squares	F-Statistic
Due to $\boldsymbol{\alpha}$ adjusted for μ and $\boldsymbol{\beta}$	$a-1$	$SS(\boldsymbol{\alpha}\vert\mu,\boldsymbol{\beta}) = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$	$F_1 = \frac{SS(\boldsymbol{\alpha}\vert\mu,\boldsymbol{\beta})/(a-1)}{SSE/(a-1)(b-1)}$
Due to $\boldsymbol{\beta}$ adjusted for μ and $\boldsymbol{\alpha}$	$b-1$	$SS(\boldsymbol{\beta}\vert\mu,\boldsymbol{\alpha}) = \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$	$F_2 = \frac{SS(\boldsymbol{\beta}\vert\mu,\boldsymbol{\alpha})/(b-1)}{SSE/(a-1)(b-1)}$
Error	$(a-1)(b-1)$	$SSE = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$	
Total	$ab-1$	$SST = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{ab}$	