4.4.2 Estimability: estimable functions

Basically, the rough idea of an estimable function is a linear function of the parameters for which an estimator can be found from $\boldsymbol{\beta}^0$ that is invariant to whatever solution to the normal equations that is used for $\boldsymbol{\beta}^0$. There are a number of examples of this kind. In this section, we confine ourselves to linear functions of the form $\boldsymbol{q}^\top\boldsymbol{\beta}$ where $\boldsymbol{q}^\top$ is a row vector.

(a) Definition: Estimable Functions

A linear function of the parameters is defined as *estimable* if it is identically equal to some linear function of the expected value of the vector of observations. This means that $\boldsymbol{q}^\top\boldsymbol{\beta}$ is estimable if $\boldsymbol{q}^\top\boldsymbol{\beta} = \boldsymbol{t}^\top E(\boldsymbol{y})$ for some vector $\boldsymbol{t}$. In other words, if a vector $\boldsymbol{t}$ exists such that $\boldsymbol{t}^\top E(\boldsymbol{y}) = \boldsymbol{q}^\top\boldsymbol{\beta}$, then $\boldsymbol{q}^\top\boldsymbol{\beta}$ is said to be estimable.

(b) Properties of Estimable Functions

(i) *The expected value of any observation is estimable.* The definition of an estimable function is that $\boldsymbol{q}^\top\boldsymbol{\beta} = \boldsymbol{t}^\top E(\boldsymbol{y})$ for some vector $\boldsymbol{t}$. Consider a $\boldsymbol{t}$ which has one element unity and the others zero. Then, $\boldsymbol{t}^\top E(\boldsymbol{y})$ will be estimable. It is an element of $E(\boldsymbol{y})$, the expected value of an observation. For example, for the linear model in the motivating example,

$$E(y_{1j}) = \mu + \alpha_1$$

and so $\mu + \alpha_1$ is estimable.

(ii) *Linear combinations of estimable functions are estimable.* Every estimable function is a linear combination of the elements of $E(\boldsymbol{y})$. This is also true about a linear combination of estimable functions. Thus, linear combinations of estimable functions are estimable. More formally, if $\boldsymbol{q_1}^\top\boldsymbol{\beta}$ and $\boldsymbol{q_2}^\top\boldsymbol{\beta}$ are estimable, there exists $\boldsymbol{t}_1^\top$ and $\boldsymbol{t}_2^\top$ such that $\boldsymbol{q_1}^\top\boldsymbol{\beta} = \boldsymbol{t}_1^\top E(\boldsymbol{y})$ and $\boldsymbol{q_2}^\top\boldsymbol{\beta} = \boldsymbol{t}_2^\top E(\boldsymbol{y})$. Hence, a linear combination $c_1\boldsymbol{q_1}^\top\boldsymbol{\beta} + c_2\boldsymbol{q_2}^\top\boldsymbol{\beta} = (c_1\boldsymbol{t}_1^\top + c_2\boldsymbol{t}_2^\top)E(\boldsymbol{y})$ and so it is estimable.

(iii) *The form of an estimable function.* if $\boldsymbol{q}^\top\boldsymbol{\beta}$ is estimable, by its definition, we have that for some vector $\boldsymbol{t}$

$$\begin{aligned} E(\boldsymbol{t}^\top\boldsymbol{y}) &= \boldsymbol{q}^\top\boldsymbol{\beta} \qquad (1)\\ \Rightarrow \boldsymbol{t}^\top E(\boldsymbol{y}) &= \boldsymbol{q}^\top\boldsymbol{\beta}\\ \Rightarrow \boldsymbol{t}^\top\boldsymbol{X}\boldsymbol{\beta} &= \boldsymbol{q}^\top\boldsymbol{\beta}. \end{aligned}$$

Since estimability is a concept that does not depend on the value of $\boldsymbol{\beta}$, the result in (1) must be true for all $\boldsymbol{\beta}$. Therefore,

$$\boxed{\boldsymbol{t}^\top\boldsymbol{X} = \boldsymbol{q}^\top} \qquad q = X^\top t$$

for some $\boldsymbol{t}$. This is equivalent to saying that $\boldsymbol{q}$ is in the *row space* of $\boldsymbol{X}$, the vector space generated by the linear combinations of the rows of $\boldsymbol{X}$.

(iii) *Invariance to the solution* $\boldsymbol{\beta}^0$. When $\boldsymbol{q}^\top\boldsymbol{\beta}$ is estimable, $\boldsymbol{q}^\top\boldsymbol{\beta}^0$ is invariant to whatever solution of

non-full rank case estimate target $\beta^0 = H\beta$.

the normal equations. This is true because

$$\boldsymbol{q}^\top \boldsymbol{\beta^0} = \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{\beta^0} = \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{y}$$

and $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top$ is invariant to the choice of $\boldsymbol{G}$. Therefore, $\boldsymbol{q}^\top \boldsymbol{\beta^0}$ is invariant to $\boldsymbol{G}$ and hence invariant to $\boldsymbol{\beta}^0$.

(iv) *The best linear unbiased estimator (BLUE) Gauss-Markov Theorem.* In chapter 3, we established this property for the full-rank linear model. We now establish that for estimable linear combinations of the parameters, the estimable linear combinations of solutions to the normal equations are best linear unbiased estimator for the less than full-rank case.

Theorem 1. (Gauss-Markov Theorem) The best linear unbiased estimator of the estimable function $\boldsymbol{q}^\top \boldsymbol{\beta}$ is $\boldsymbol{q}^\top \boldsymbol{\beta}^0$.

Proof. We prove this theorem from the following three aspects:

(i) $\boldsymbol{q}^\top \boldsymbol{\beta}^0 = \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{y} \Longleftrightarrow$ linear function of y_i.

(ii)

scalar

linear combition of β is estimable.

$$\begin{aligned} E(\boldsymbol{q}^\top \boldsymbol{\beta}^0) &= \boldsymbol{q}^\top E(\boldsymbol{\beta}^0) \\ &= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top E(\boldsymbol{y}) \\ &= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{q}^\top \boldsymbol{\beta}. \end{aligned}$$

$\Rightarrow$ unbiased estimator.

(iii) Minimum variance. (Optimality)

$GX^\top y$.

$$\begin{aligned} \boxed{var(\boldsymbol{q}^\top \boldsymbol{\beta}^0)} &= \boldsymbol{q}^\top var(\boldsymbol{\beta}^0)\boldsymbol{q} \\ &= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{G}^\top \boldsymbol{q}\sigma^2 \\ &= \boxed{\boldsymbol{t}^\top \boldsymbol{X}}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{G}^\top \boxed{\boldsymbol{X}^\top \boldsymbol{t}}\sigma^2 \\ &= \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{t}\sigma^2 \\ &\boxed{= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{q}\sigma^2.} \end{aligned}$$

Suppose $\boldsymbol{k}^\top \boldsymbol{y}$ is another linear unbiased estimator of $\boldsymbol{q}^\top \boldsymbol{\beta}$ different from $\boldsymbol{q}^\top \boldsymbol{\beta}^0$.

$$\begin{aligned} &\Rightarrow E(\boldsymbol{k}^\top \boldsymbol{y}) = \boldsymbol{q}^\top \boldsymbol{\beta} \\ &\Rightarrow \boldsymbol{k}^\top \boldsymbol{X} = \boldsymbol{q}^\top \\ &\Rightarrow \boxed{cov(\boldsymbol{q}^\top \boldsymbol{\beta}^0, \boldsymbol{k}^\top \boldsymbol{y})} \\ &= cov(\underline{\boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top} \boldsymbol{y}, \boldsymbol{k}^\top \boldsymbol{y}) \\ &= \underline{\boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top}(\boldsymbol{I}\sigma^2)\boldsymbol{k} \\ &\boxed{= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{q}\sigma^2.} \end{aligned}$$

$q^\top \beta^0$ has a invariant covariance with any other unbiased estimator including itself. $q^\top G q \sigma^2$.

Now,

$$
\begin{aligned}
& var(\boldsymbol{q}^\top \boldsymbol{\beta}^0 - \boldsymbol{k}^\top \boldsymbol{y}) \\
=& var(\boldsymbol{q}^\top \boldsymbol{\beta}^0) + var(\boldsymbol{k}^\top \boldsymbol{y}) - 2cov(\boldsymbol{q}^\top \boldsymbol{\beta}^0, \boldsymbol{k}^\top \boldsymbol{y}) \\
=& var(\boldsymbol{k}^\top \boldsymbol{y}) + \boxed{\boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{q}\sigma^2 - 2\boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{q}\sigma^2.} \\
\Rightarrow \quad & var(\boldsymbol{q}^\top \boldsymbol{\beta}^0 - \boldsymbol{k}^\top \boldsymbol{y}) \\
=& var(\boldsymbol{k}^\top \boldsymbol{y}) - var(\boldsymbol{q}^\top \boldsymbol{\beta}^0) \geq 0 \\
\Rightarrow \quad & var(\boldsymbol{k}^\top \boldsymbol{y}) \geq var(\boldsymbol{q}^\top \boldsymbol{\beta}^0) \\
\Rightarrow \quad & \boldsymbol{q}^\top \boldsymbol{\beta}^0 \text{ is B.L.U.E. of } \boldsymbol{q}^\top \boldsymbol{\beta}.
\end{aligned}
$$

Remark: Under normality assumption, $\boldsymbol{q}^\top \boldsymbol{\beta}^0 \sim N(\boldsymbol{q}^\top \boldsymbol{\beta}, \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{q}\sigma^2)$.

4.4.3 Test of Estimability

q is in the row space of X. which means q is a special data

A given function $\boldsymbol{q}^\top \boldsymbol{\beta}$ is estimable if some vector $\boldsymbol{t}$ can be found such that $\boldsymbol{t}^\top \boldsymbol{X} = \boldsymbol{q}^\top$. However, for $\boldsymbol{q}^\top$ known, derivation of a $\boldsymbol{t}^\top$ satisfying $\boldsymbol{t}^\top \boldsymbol{X} = \boldsymbol{q}^\top$ may not always be easy especially if $\boldsymbol{X}$ has large dimensions. Instead of finding $\boldsymbol{t}^\top$, it can be determined whether $\boldsymbol{q}^\top \boldsymbol{\beta}$ is estimable by checking whether $\boldsymbol{q}^\top$ satisfies the equation $\boldsymbol{q}^\top \boldsymbol{H} = \boldsymbol{q}^\top$, where $\boldsymbol{H} = \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}$. We state this result in Theorem 2 below.

Theorem 2. The linear function $\boldsymbol{q}^\top \boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{q}^\top \boldsymbol{H} = \boldsymbol{q}^\top$ where $\boldsymbol{H} = \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}$.

$t^\top X = q^\top \quad \Leftrightarrow \quad q^\top H = q^\top$.

Proof. If $\boldsymbol{q}^\top \boldsymbol{\beta}$ is estimable, there exists a vector $\boldsymbol{t}$ such that $\boldsymbol{t}^\top \boldsymbol{X} = \boldsymbol{q}^\top$,

$\Rightarrow$

$$
\begin{aligned}
\Rightarrow \quad \boldsymbol{q}^\top \boldsymbol{H} &= \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X} \\
&= \boldsymbol{t}^\top \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X} \\
&= \boldsymbol{t}^\top \boldsymbol{X} = \boldsymbol{q}^\top.
\end{aligned}
$$

On the other hand, if $\boldsymbol{q}^\top \boldsymbol{H} = \boldsymbol{q}^\top$,

$\Leftarrow$

$$
\begin{aligned}
\Rightarrow \quad & \boldsymbol{q}^\top = \boxed{\boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top} \boldsymbol{X}, \\
\Rightarrow \quad & \text{take } \boldsymbol{t}^\top = \boldsymbol{q}^\top \boldsymbol{G}\boldsymbol{X}^\top, \\
& \text{we have } \boldsymbol{q}^\top = \boldsymbol{t}^\top \boldsymbol{X}. \\
\Rightarrow \quad & \boldsymbol{q}^\top \boldsymbol{\beta} \text{ is estimable.}
\end{aligned}
$$

Example. Consider the normal equations

$$
(\boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y},
$$

where $(\boldsymbol{X}^\top \boldsymbol{X}) = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} 14 \\ 6 \\ 8 \end{pmatrix}.$

One possible generalized inverse is

$$\Rightarrow \quad \boldsymbol{G} = (\boldsymbol{X}^\top \boldsymbol{X})^- = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$\begin{aligned} \boldsymbol{\beta}^0 = \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{y} &= \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 14 \\ 6 \\ 8 \end{pmatrix} \\ &= \begin{pmatrix} 8 \\ -2 \\ 0 \end{pmatrix}. \end{aligned}$$

Another generalized inverse is

$$\boldsymbol{G}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{\beta}_1^0 = \boldsymbol{G_1}\boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 14 \\ 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \\ 8 \end{pmatrix}.$$

The matrix

$$\underline{\boldsymbol{H}} = \boldsymbol{G_1}\boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

depend on the choice of G.

QE

- Is $\beta_1 - \beta_2$ estimable?

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\boldsymbol{q}^\top = (0\ 1\ \ -1)$$

$$\Rightarrow \quad \boldsymbol{q}^\top \boldsymbol{H} = (0\ 1\ \ -1) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = (0\ 1\ \ -1) = \boldsymbol{q}^\top$$

$$\Rightarrow \quad \boldsymbol{q}^\top \boldsymbol{\beta} \text{ is estimable.}$$

- Is $\beta_1 + \beta_2$ estimable?

$$\begin{aligned}
\boldsymbol{q}^\top &= (0\ 1\ 1) \\
\boldsymbol{q}^\top \boldsymbol{H} &= (0\ 1\ 1) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \\
&= (2\ 1\ 1) \neq \boldsymbol{q}^\top. \\
\Rightarrow \quad & \beta_1 + \beta_2 \text{ is not estimable.}
\end{aligned}$$

- Is $3\beta_0 - \beta_1 - 2\beta_2$ estimable?

$$\begin{aligned}
\boldsymbol{q}^\top &= (3\ \ -1\ \ -2) \\
\boldsymbol{q}^\top \boldsymbol{H} &= (3\ \ -1\ \ -2) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \\
&= (-3\ \ -1\ \ -2) \neq \boldsymbol{q}^\top \\
\Rightarrow \quad & 3\beta_0 - \beta_1 - 2\beta_2 \text{ is not estimable.}
\end{aligned}$$

4.5 The general linear hypothesis

In chapter 3, we developed the theory for testing the general linear hypothesis $H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$ for the full-rank case. We shall now develop this theory for the non-full-rank case. In the non-full-rank case, we can test some hypothesis. Others, we cannot. We shall establish conditions for "testability" of a hypothesis.

4.5.1 Testable hypothesis

A testable hypothesis is one that can be expressed in terms of estimable functions. It seems reasonable that a testable hypothesis should be made up of estimable functions because the results for the full-rank case suggest that $\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m}$ will be part of the test statistic. If this is the case, $\boldsymbol{K}^\top \boldsymbol{\beta}^{\mathbf{0}}$ will need to be invariant to $\boldsymbol{\beta}^0$. This can only happen if $\boldsymbol{K}^\top \boldsymbol{\beta}$ consists of estimable functions.

In light of the above considerations, a testable hypothesis $H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \underset{\in\mathbb{R}^r}{\boldsymbol{m}}$ is taken as one, where $\boldsymbol{K}^\top \boldsymbol{\beta} \equiv \{\boldsymbol{k}_i^\top \boldsymbol{\beta}\}$ for $i = 1, \ldots, r$ such that $\boldsymbol{k}_i^\top \boldsymbol{\beta}$ is estimable for all i. Hence $\boldsymbol{k}_i^\top = \boldsymbol{t}_i^\top X$ for some $\boldsymbol{t}_i^\top$. As a result,

$$\boxed{\boldsymbol{K}^\top = \boldsymbol{T}^\top \boldsymbol{X}}$$

for some matrix $\boldsymbol{T}^\top$ of order $r \times n$. Furthermore, any hypothesis is considered only in terms of its linearly independent components. Therefore, the matrix $\boldsymbol{K}^\top$ of size $r \times p$ is always of full-row rank.

Since $\boldsymbol{K}^\top \boldsymbol{\beta}$ is taken to be a set of estimable functions their BLUE are

$$\widehat{\boldsymbol{K}^\top \boldsymbol{\beta}} = \boldsymbol{K}^\top \boldsymbol{\beta}^{\mathbf{0}}$$

with expectation

$$E(\boldsymbol{K}^\top \boldsymbol{\beta}^{\mathbf{0}}) = \boldsymbol{K}^\top \boldsymbol{\beta}.$$

To show $\operatorname{rank}(K^\top G K) = \operatorname{rank}(K^\top) = r$

then $r(K^\top) = r(K^\top G X^\top) \geq r(K^\top G X^\top X) = r(K^\top) = r$

Since $K^\top\beta$ is a set of estimable function

$K^\top H = K^\top \Rightarrow K^\top G X^\top X = K^\top$

thus

$r = r(K^\top G X^\top (K^\top G X^\top)^\top)$

$= r(K^\top G \underbrace{X^\top X G^\top K}_{H^\top K = K.})$

$= r(K^\top G K)$.

The BLUE's have variance

$\beta^0 = G X^\top y$

$$
\begin{aligned}
var(\widehat{\boldsymbol{K}^\top \boldsymbol{\beta}}) &= \boldsymbol{K}^\top var(\boldsymbol{\beta}^0) \boldsymbol{K} \\
&= \boldsymbol{K}^\top \boldsymbol{G} \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{G}^\top \boldsymbol{K} \sigma^2 \\
&= \boldsymbol{K}^\top \boldsymbol{G} \underbrace{\boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{G}^\top \boldsymbol{X}^\top}_{X^\top} \underbrace{\boldsymbol{T}}_{K} \sigma^2 \\
&= \boldsymbol{K}^\top \boldsymbol{G} \boldsymbol{K} \sigma^2.
\end{aligned}
$$

A key step is to show $\boldsymbol{K}^\top \boldsymbol{G} \boldsymbol{K}$ is non-singular. (show on the board)

to show $\operatorname{rank}(K^\top G K) = r$

4.5.2 Hypothesis testing

The test for the testable hypothesis $H_0 :\ \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$ (Let the number of rows in $\boldsymbol{K}^\top = r$) is developed just as in the full-rank case. In addition, we assume that $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\sigma^2 I})$. And hence

$$\boldsymbol{\beta}^0 \sim N(\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{G}^\top \boldsymbol{\sigma^2})$$

and

$$\boldsymbol{K}^\top \boldsymbol{\beta}^0 - \boldsymbol{m} \sim N(\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m}, \boldsymbol{K}^\top \boldsymbol{G}\boldsymbol{K}\sigma^2).$$

Take

$$Q = (\boldsymbol{K}^\top \boldsymbol{\beta}^0 - m)^\top (\boldsymbol{K}^\top \boldsymbol{G}\boldsymbol{K})^{-1} (\boldsymbol{K}^\top \boldsymbol{\beta}^0 - m),$$

then

$$\frac{Q}{\sigma^2} \sim \chi^2_{r,(\boldsymbol{K}^\top \boldsymbol{\beta} - m)^\top (\boldsymbol{K}^\top \boldsymbol{G}\boldsymbol{K})^{-1} (\boldsymbol{K}^\top \boldsymbol{\beta} - m)/2\sigma^2}.$$

It is straightforward to show that

$$F(H) = \frac{Q/r}{SSE/(n - r(\boldsymbol{X}))} \sim F_{r,n-r(\boldsymbol{X}),(\boldsymbol{K}^\top \boldsymbol{\beta} - m)^\top (\boldsymbol{K}^\top \boldsymbol{G}\boldsymbol{K})^{-1} (\boldsymbol{K}^\top \boldsymbol{\beta} - m)/2\sigma^2}.$$

Under $H_0 :\ \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$,

$$F(H) = \frac{Q/r}{SSE/(n - r(\boldsymbol{X}))} \sim F_{r,n-r(\boldsymbol{X})}.$$

Furthermore, under $H_0 :\ \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$,

$$\begin{aligned} \boldsymbol{\beta}^0_H &= \boldsymbol{\beta}^0 - \boldsymbol{G}\boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{G}\boldsymbol{K})^{-1} (\boldsymbol{K}^\top \boldsymbol{\beta}^0 - \boldsymbol{m}) \\ SSE_H &= SSE + Q. \end{aligned}$$

where SSE=$\boldsymbol{y}^\top (\boldsymbol{I} - \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top)\boldsymbol{y}$. The above result is obtained by replacing $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ in the full-rank case by $\boldsymbol{G}$.

4.6 One-Way Model (Balanced design)

To summarize, the model of concern is $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \cdots, a; j = 1, \cdots, n.$

In matrix notation: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 \boldsymbol{I})$, $\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ \vdots \\ y_{a1} \\ y_{a2} \\ \vdots \\ y_{an} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \vdots \\ \varepsilon_{a1} \\ \varepsilon_{a2} \\ \vdots \\ \varepsilon_{an} \end{pmatrix},$

$$\boldsymbol{X} = \begin{array}{c} \begin{matrix} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_a \end{matrix} \\ \begin{pmatrix} 1 & 1 & 0 & & 0 \\ 1 & 1 & 0 & & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & 0 & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 1 & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & & 1 \end{pmatrix} \end{array}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} an & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ n & 0 & 0 & \cdots & n \end{pmatrix}.$$

A generalized inverse of $\boldsymbol{X}^\top \boldsymbol{X}$ is

$$\begin{aligned} \boldsymbol{G} = (\boldsymbol{X}^\top \boldsymbol{X})^- &= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{n} & \cdots & 0 \\ \vdots & 0 & \frac{1}{n} & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n} \end{pmatrix} \\ &= \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{D}\{\frac{1}{n}\} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
\boldsymbol{H} &= \boldsymbol{G}\boldsymbol{X}^\top\boldsymbol{X} \\
&= \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{D}\{\frac{1}{n}\} \end{pmatrix} \begin{pmatrix} n & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & 0 & 0 & \cdots & n \end{pmatrix} \\
&= \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{1}_a & \boldsymbol{I}_a \end{pmatrix}.
\end{aligned}$$

And

$$\boldsymbol{X}^\top\boldsymbol{y} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{pmatrix},$$

where $y_{..} \triangleq \sum_{i=1}^a \sum_{j=1}^n y_{ij}$, $y_{i.} \triangleq \sum_{j=1}^n y_{ij}$ and $\bar{y}_{i.} \triangleq (1/n)\sum_{j=1}^n y_{ij}$, $i = 1, \ldots, a$.

$$\boldsymbol{\beta}^0 = \boldsymbol{G}\boldsymbol{X}^\top\boldsymbol{y} = \begin{pmatrix} 0 \\ \bar{y}_{1.} \\ \bar{y}_{2.} \\ \vdots \\ \bar{y}_{a.} \end{pmatrix}.$$

<u>4.6.1 Analysis of Variance</u>

1.

$$\begin{aligned}
SSR &= \boldsymbol{\beta}^{0\top}\boldsymbol{X}^\top\boldsymbol{y} \\
&= (0 \ \bar{y}_{..} \ \bar{y}_{1.} \ \bar{y}_{2.} \ \cdots \ \bar{y}_{a.}) \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{pmatrix} \\
&= \sum_{i=1}^a \frac{y_{i.}^2}{n}.
\end{aligned}$$

2.
$$SSM = \frac{\boldsymbol{y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{y}}{an} = an\bar{y}_{..}^2.$$

3.
$$SSR_m = SSR - SSM = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - an\bar{y}_{..}^2.$$

4.
$$SST = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2.$$

5.
$$SST_m = SST - SSM = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - an\bar{y}_{..}^2.$$

6.
$$\begin{aligned} SSE &= SST - SSR \\ &= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \sum_{i=1}^{a} \frac{y_{i.}^2}{n}. \end{aligned}$$

Example 1. Three different treatment methods for removing organic carbon from tar sand wastewater are to be compared. The method are air otation (AF), foam seperation (FS), and ferric - chloride coagulation (FCC). The data are given as follows

AF(I)	FS(II)	FCC(III)
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0
35.8	40.9	27.1
36.1	41.0	27.5
36.5	43.2	28.1
36.8	44.9	28.1
37.2	46.9	28.7
37.4	51.6	30.7
37.7	53.6	31.2

Model:
$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon \sim N(0, \sigma^2).$$

α_i : Effect of treatment i, $\quad i = 1, 2, 3;\ j = 1, 2, \cdots, 10.$

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

$$\boldsymbol{X}^\top\boldsymbol{X} = \begin{pmatrix} 30 & 10 & 10 & 10 \\ 10 & 10 & 0 & 0 \\ 10 & 0 & 10 & 0 \\ 10 & 0 & 0 & 10 \end{pmatrix}.$$

A generalized inverse of $\boldsymbol{X}^\top\boldsymbol{X}$ is

$$(\boldsymbol{X}^\top\boldsymbol{X})^- = \begin{pmatrix} 0 & 0 & 0 & 10 \\ 0 & \frac{1}{10} & 0 & 0 \\ 0 & 0 & \frac{1}{10} & 0 \\ 0 & 0 & 0 & \frac{1}{10} \end{pmatrix}.$$

$$\boldsymbol{\beta}^0 = (\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{X}^\top\boldsymbol{y} = \begin{pmatrix} 0 \\ 36.25 \\ 44 \\ 28.18 \end{pmatrix}.$$

SST= 40720.41,

SSR= 40441.749,

SSE= 278.661,

MSE= $\frac{278.661}{27} = 10.32 = \hat{\sigma}^2$.

Hypothesis testing of treatment effects.

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3,$$

with reference to section 4.5.2,

$$\boldsymbol{K}^\top = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \qquad \boldsymbol{m} = 0.$$

$$\begin{aligned} &\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{K} \\ =& \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{10} & 0 & 0 \\ 0 & 0 & \frac{1}{10} & 0 \\ 0 & 0 & 0 & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \\ =& \frac{1}{10} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

$$(\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{K})^{-1} = \frac{1}{10} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

The test statistics is

$$\begin{aligned} F(H) &= \frac{Q/r}{\hat{\sigma}^2} \\ &= \frac{(\boldsymbol{K}^\top \hat{\boldsymbol{\beta}}^{\mathbf{0}} - \boldsymbol{m})^\top (\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X}) \boldsymbol{K})^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}}^{\mathbf{0}} - \boldsymbol{m})/2}{10.32} \\ &= 60.63. \qquad (Q = 1251.533) \end{aligned}$$

Since p-value<0.01, there is strong evidence to reject H_0.

Anova Table

Source of Variance	SS	df	MS	F
Regression(Full)	40441.749	3		
Reduced Model	39190.216	1		
Hypothesis	1251.533	2	625.7665	60.63
Residual	278.661	27	10.32	
Total	40720.41	30		

4.6.2 The"Usual Constraints"

A source of difficulty with a non-full-rank model is that the normal equation do not have a unique solution. We have skirted this situation by using a generalized inverse of $\boldsymbol{X}^\top \boldsymbol{X}$. Another way to obtain a solution to the normal equation is to impose the "usual constraints" or usual restrictions. A common constraint or side condition is

$$\sum_{i=1}^{a} \alpha_i^* = 0.$$

Denote the constrained parameters as $\mu^*, \alpha_1^*, \cdots, \alpha_a^*$. With the reparametrization, the model becomes

$$y_{ij} = \mu_i + \epsilon_{ij} = \mu^* + \alpha_i^* + \epsilon_{ij}, \tag{2}$$

where $\sum_{i=1}^{a} \alpha_i^* = 0$, $E(\epsilon_{ij}) = 0$, $Var(\epsilon_{ij}) = \sigma^2$ for all i, j, $Cov(\epsilon_{ij}, \epsilon_{rs}) = 0$ for any $(i,j) \neq (r,s)$. In addition, $\mu_i = \mu + \alpha_i$ is the effect of treatment i.

The normal equation is

$$\begin{pmatrix} an & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & 0 & 0 & \cdots & n \end{pmatrix} \begin{pmatrix} \mu^* \\ \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_a^* \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{pmatrix}. \tag{3}$$

With the side condition $\sum_{i=1}^{a} \alpha_i^* = 0$, we have

$\hat{\mu^*} = \frac{y_{..}}{an} = \bar{y}_{..}$, (where $\bar{y}_{..}$ is the overall mean) and $\hat{\alpha}_i^* = \bar{y}_{i.} - \bar{y}_{..}$ for $i = 1, \cdots, a$, that is

$$\hat{\beta} = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{a.} - \bar{y}_{..} \end{pmatrix}.$$

Next, we consider the test of equality of means as follows:

$H_0: \ \mu_1 = \cdots = \mu_a \Longleftrightarrow$

$H_0: \ \alpha_1 = \cdots = \alpha_a$ (this is testable, as it can be expressed in terms of $(k-1)$ linearly independent estimable functions.) $\Longleftrightarrow$

$H_0: \ \alpha_1^* = \cdots = \alpha_a^*$ (subject to the side condition $\sum_{i=1}^a \alpha_i^* = 0$) $\Longleftrightarrow$

$H_0: \ \alpha_1^* = \cdots = \alpha_a^* = 0$.

To this end, we first consider the full model (2):

$$\begin{aligned} SS(\mu, \alpha) &= \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{y} \\ &= (\bar{y}_{..}, \bar{y}_{1.} - \bar{y}_{..}, \cdots, \bar{y}_{a.} - \bar{y}_{..}) \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{pmatrix} \\ &= \sum_{i=1}^a \frac{y_{i.}^2}{n}. \end{aligned}$$

Next, we consider the reduced model under H_0:

$$SS(\mu) = \frac{y_{..}^2}{an},$$

which is the SSM for the full model when assuming there is only intercept term in the model. Hence, the sum of squares for the α's adjusted for μ is given by

$$SS(\alpha|\mu) = SS(\mu, \alpha) - SS(\mu) = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2.$$

The following is the ANOVA table for testing $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a$ in the One-way model:

Anova Table

Source	df	Sum of Squares	Mean Square	F
Treatment	$a-1$	$SS(\alpha\|\mu) = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$	$\frac{SS(\mu\|\alpha)}{a-1}$	$\frac{SS(\mu\|\alpha)/(a-1)}{SSE/a(n-1)}$
Error	$a(n-1)$	$SSE = \sum_j \sum_i y_{ij}^2 - \sum_{i=1}^a \frac{y_{i.}^2}{n}$	$\frac{SSE}{a(n-1)}$	
Total	$an-1$	$SST_m = \sum_j \sum_i y_{ij}^2 - \frac{y_{..}^2}{an}$		

Due to limitation of time, I will leave other sections on orthogonal contrasts in the One-way model models, two-way models and two-way models with interaction terms for students' self-learning with reference to the textbook Searle and Gruber (2017).