13.1. Testable Hypotheses

Suppose we want to test the hypothesis:

$$H : \mathbf{A}_{q\times p}\boldsymbol{\beta}_{p\times 1} = \mathbf{0}_{q\times 1}.$$

In terms of the rows of $\mathbf{A}$ this can be written as

$$\begin{pmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_q' \end{pmatrix} \boldsymbol{\beta} = \mathbf{0}$$

i.e., $\mathbf{a}_i'\boldsymbol{\beta} = 0$ for each row of $\mathbf{A}$.

Definition: The hypothesis $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is *testable* if $\mathbf{a}_i'\boldsymbol{\beta}$ is an estimable function for each row $\mathbf{a}_i$ of $\mathbf{A}$.

Note: Recall that $\mathbf{a}_i'\boldsymbol{\beta}$ is estimable if $\mathbf{a}_i' = \mathbf{b}_i'\mathbf{X}$ for some $\mathbf{b}_i$. Therefore $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is testable if $\mathbf{A} = \mathbf{M}\mathbf{X}$ for some $\mathbf{M}$, i.e., the rows of $\mathbf{A}$ are linearly combinations of the rows of $\mathbf{X}$.

Example: (One-way ANOVA with 3 groups).

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \\ Y_{31} \\ \vdots \\ Y_{3J} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1J} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2J} \\ \varepsilon_{31} \\ \vdots \\ \varepsilon_{3J} \end{pmatrix}$$

Examples of testable hypotheses are:

1. $H : (1, 1, 0, 0)\boldsymbol{\beta} = \mu + \alpha_1 = 0$
2. $H : (1, 0, 1, 0)\boldsymbol{\beta} = \mu + \alpha_2 = 0$
3. $H : (1, 0, 0, 1)\boldsymbol{\beta} = \mu + \alpha_3 = 0$
4. $H : (0, 1, -1, 0)\boldsymbol{\beta} = \alpha_1 - \alpha_2 = 0$
5. $H : \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \boldsymbol{\beta} = \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,
   i.e., $\alpha_1 = \alpha_2 = \alpha_3$ (no group effects).

13.2. Development of the F-Test

How can we test $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$?

The idea is to fit the model with and without the linear restrictions. The RSS in the restricted case will always be larger. However, if the hypothesis is true, the "additional" reduction in the RSS from fitting the unrestricted model will be small – only due to "chance." Said differently: if the hypothesis H is false, than the reduction in RSS from fitting the unrestricted model will be more than we would expect by chance.

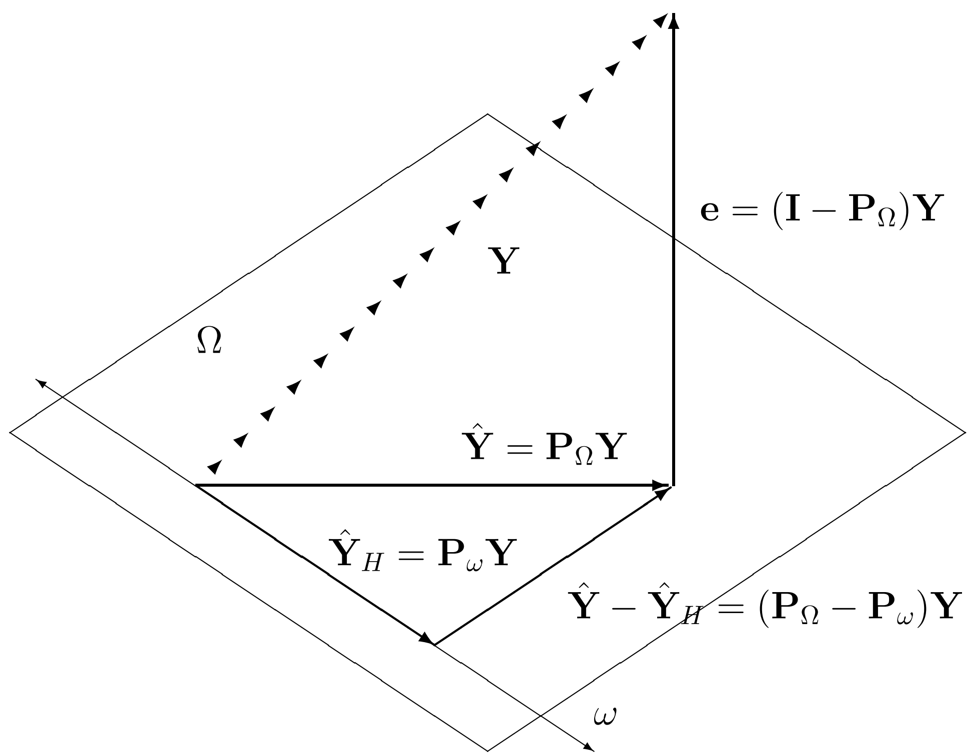
We compare the residual sum of squares (RSS) for the full model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ to the residual sum of squares (RSS_H) for the restricted model (with $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$).

Let $\boldsymbol{\mu} = E[\mathbf{Y}]$. Under the full model, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \in \mathcal{R}(\mathbf{X}) \equiv \Omega$. If $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\mathbf{A} = \mathbf{M}\mathbf{X}$, then

$$\begin{aligned} H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0} &\Leftrightarrow H : \mathbf{M}\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \\ &\Leftrightarrow H : \mathbf{M}\boldsymbol{\mu} = \mathbf{0} \\ &\Leftrightarrow H : \boldsymbol{\mu} \in \mathcal{R}(\mathbf{X}) \cap \mathcal{N}(\mathbf{M}) \equiv \omega. \end{aligned}$$

(Recall: $\mathcal{N}(\mathbf{M}) = \{\mathbf{u} : \mathbf{M}\mathbf{u} = \mathbf{0}\}$) is the null space of $\mathbf{M}$.)

Thus we have translated a hypothesis about $\boldsymbol{\beta}$ into a hypothesis about $\boldsymbol{\mu} = E[\mathbf{Y}]$. We can write ω as $\{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \mathbf{A}\boldsymbol{\beta} = \mathbf{0}\}$ or, equivalently, $\omega = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \mathbf{M}\boldsymbol{\mu} = \mathbf{0}\}$. (see figure on following page)


$\mathbf{e} = (\mathbf{I} - \mathbf{P}_\Omega)\mathbf{Y}$
$\mathbf{Y}$
Ω
$\hat{\mathbf{Y}} = \mathbf{P}_\Omega \mathbf{Y}$
$\hat{\mathbf{Y}}_H = \mathbf{P}_\omega \mathbf{Y}$
$\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_H = (\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y}$
ω

Let $\hat{\mathbf{Y}} = \mathbf{P}_\Omega \mathbf{Y}$ and $\hat{\mathbf{Y}}_H = \mathbf{P}_\omega \mathbf{Y}$ be the orthogonal projections of $\mathbf{Y}$ onto Ω and ω. The RSS for the full model is

$$\begin{aligned} RSS &= (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) = (\mathbf{Y} - \mathbf{P}_\Omega \mathbf{Y})'(\mathbf{Y} - \mathbf{P}_\Omega \mathbf{Y}) \\ &= ((\mathbf{I} - \mathbf{P}_\Omega)\mathbf{Y})'((\mathbf{I} - \mathbf{P}_\Omega)\mathbf{Y}) = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{Y}. \end{aligned}$$

Similarly, the RSS for the restricted model (with $\boldsymbol{\mu} \in \omega$) is

$$RSS_H = (\mathbf{Y} - \hat{\mathbf{Y}}_H)'(\mathbf{Y} - \hat{\mathbf{Y}}_H) = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_\omega)\mathbf{Y}.$$

Therefore,

$$RSS_H - RSS = \mathbf{Y}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y}.$$

Note: All of the above is linear algebra. We haven't yet used anything about the hypothesis H other than the fact that it corresponds to a subspace ω.

Now, since $\boldsymbol{\mu} \in \Omega$, we can continue:

$$RSS = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{Y} = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{I} - \mathbf{P}_\Omega)(\mathbf{Y} - \boldsymbol{\mu})$$

since $(\mathbf{I} - \mathbf{P}_\Omega)\boldsymbol{\mu} = \mathbf{0}$. Furthermore, <u>if H is true</u> then $\boldsymbol{\mu} \in \omega$ and we have

$$RSS_H = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_\omega)\mathbf{Y} = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{I} - \mathbf{P}_\omega)(\mathbf{Y} - \boldsymbol{\mu}).$$

and thus

$$RSS_H - RSS = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{P}_\Omega - \mathbf{P}_\omega)(\mathbf{Y} - \boldsymbol{\mu}).$$

13.3. Distribution of the F-Statistic when H is true

Definition: Let X_1 and X_2 be independent random variables with $X_1 \sim \chi^2_{d_1}$ and $X_2 \sim \chi^2_{d_2}$. Then the distribution of the ratio

$$F = \frac{X_1/d_1}{X_2/d_2}$$

is the F ***distribution*** with d_1 numerator degrees of freedom (df) and d_2 denominator df and is denoted F_{d_1,d_2}.

Theorem 13.3.1: If $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\text{rank}(\mathbf{A}_{q\times p}) = q$, then, when H is true,

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-r)} \sim F_{q,n-r},$$

the F distribution with q and $n-r$ degrees of freedom.

Proof: We just derived that

$$RSS/\sigma^2 = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{I} - \mathbf{P}_\Omega)(\mathbf{Y} - \boldsymbol{\mu})/\sigma^2.$$

Similarly, if $\boldsymbol{\mu} \in \omega$,

$$RSS_H/\sigma^2 = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{I} - \mathbf{P}_\omega)(\mathbf{Y} - \boldsymbol{\mu})/\sigma^2,$$

and hence

$$(RSS_H - RSS)/\sigma^2 = (\mathbf{Y} - \boldsymbol{\mu})'(\mathbf{P}_\Omega - \mathbf{P}_\omega)(\mathbf{Y} - \boldsymbol{\mu})/\sigma^2.$$

Because $\text{rank}(\mathbf{I} - \mathbf{P}_\Omega) = n - r$,

$$RSS/\sigma^2 \sim \chi^2_{n-r}.$$

Also

$$(RSS_H - RSS)/\sigma^2 \sim \chi^2_q,$$

when the null hypothesis $H : \boldsymbol{\mu} \in \omega$ is true.

Now we have the following decomposition of χ^2 variables:

$$Q_1 = Q + Q_2$$

where

$$Q_2 \equiv RSS/\sigma^2 \sim \chi^2_{n-r},$$
$$Q \equiv (RSS_H - RSS)/\sigma^2 \sim \chi^2_q,$$

and, by addition of the degrees of freedom,

$$Q_1 = RSS_H/\sigma^2 \sim \chi^2_{n-(r-q)}.$$

By Hogg & Craig Theorem (Lecture 6, page 8), $Q_1 - Q_2 = Q$ and Q_2 are independent. Therefore the F ratio has the distribution of a ratio of independent χ^2 variables divided by their *df*, which is the definition of the F distribution.

13.4. Change in RSS

The following theorem gives a useful expression for the change in residual sum of squares.

Theorem 13.4.1: If $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis, then

$$RSS_H - RSS = (\mathbf{A}\hat{\boldsymbol{\beta}})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-}(\mathbf{A}\hat{\boldsymbol{\beta}}).$$

To prove this we need a lemma:

Lemma: Let $\Omega = \mathcal{R}(\mathbf{X})$ and $\omega = \Omega \cap \mathcal{N}(\mathbf{M})$. Then

1. $\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{P}_{\omega^\perp \cap \Omega}$ (Seber & Lee, B3.2).
2. $\omega^\perp \cap \Omega = \mathcal{R}(\mathbf{P}_\Omega \mathbf{M}')$ (Seber & Lee, B3.3).
3. If $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis,
   $\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-}\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$

Proof of Part 3 of the Lemma, which says
$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^-\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$:

Parts 1 and 2 of the Lemma tell us that $\mathbf{P}_\Omega - \mathbf{P}_\omega$ is equal to $\mathbf{P}_{\mathcal{R}(\mathbf{P}_\Omega\mathbf{M}')}$.

For any matrix $\mathbf{X}$, $\mathbf{P}_{\mathcal{R}(\mathbf{X})} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ defines projection onto $\mathcal{R}(\mathbf{X})$. In particular, $\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$, and for the matrix $\mathbf{P}_\Omega\mathbf{M}'$

$$\begin{aligned}\mathbf{P}_{\mathcal{R}(\mathbf{P}_\Omega\mathbf{M}')} &= (\mathbf{P}_\Omega\mathbf{M}')[(\mathbf{P}_\Omega\mathbf{M}')'(\mathbf{P}_\Omega\mathbf{M}')]^-(\mathbf{P}_\Omega\mathbf{M}')' \\ &= \mathbf{P}_\Omega\mathbf{M}'[\mathbf{M}\mathbf{P}_\Omega\mathbf{M}']^-\mathbf{M}\mathbf{P}_\Omega.\end{aligned}$$

Using $\mathbf{A} = \mathbf{M}\mathbf{X}$,

$$\mathbf{P}_\Omega\mathbf{M}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{M}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{A}',$$

$$\mathbf{M}\mathbf{P}_\Omega = \mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{X}',$$

$$\mathbf{M}\mathbf{P}_\Omega\mathbf{M}' = \mathbf{M}(\mathbf{P}_\Omega\mathbf{M}') = \underbrace{\mathbf{M}\mathbf{X}}_{\mathbf{A}}(\mathbf{X}'\mathbf{X})^-\mathbf{A}' = \mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}',$$

Hence

$$\begin{aligned}\mathbf{P}_{\mathcal{R}(\mathbf{P}_\Omega\mathbf{M}')} &= (\mathbf{P}_\Omega\mathbf{M}')([\mathbf{M}\mathbf{P}_\Omega\mathbf{M}']^-)(\mathbf{M}\mathbf{P}_\Omega) \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^-\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\end{aligned}$$

Proof of Theorem 13.4.1:

$$\begin{aligned} RSS_H - RSS &= \mathbf{Y}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^-\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{Y} \\ &= \hat{\boldsymbol{\beta}}'\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^-\mathbf{A}\hat{\boldsymbol{\beta}} \\ &= (\mathbf{A}\hat{\boldsymbol{\beta}})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^-(\mathbf{A}\hat{\boldsymbol{\beta}}) \end{aligned}$$

This result shows that large values of $RSS_H - RSS$ indicate departures from $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$.

If $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, then $\mathbf{A}\hat{\boldsymbol{\beta}}$ will tend to be close to 0 and so $RSS_H - RSS$ will tend to be small.

Note that this is just a quadratic form for the random vector $\mathbf{A}\hat{\boldsymbol{\beta}}$. Letting rank$(\mathbf{A}) = q$, we can derive :

$$E[RSS_H - RSS] = \sigma^2 q + (\mathbf{A}\boldsymbol{\beta})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^-\mathbf{A}']^{-1}(\mathbf{A}\boldsymbol{\beta}).$$

Therefore, when $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is true,

$$E[RSS_H - RSS] = \sigma^2 q.$$

We form a test statistic by making a ratio:

$$\frac{RSS_H - RSS}{q\hat{\sigma}^2} = \frac{(RSS_H - RSS)/q}{RSS/(n-r)}.$$

Under H, F will be approximately equal to 1. If H is not true F will tend to be larger than 1.