

## 11 Hypothesis Testing

### 11.1 Introduction

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  $\mathbf{A}$  (here  $\mathbf{a}'_i$  denotes the  $i$ th row of  $\mathbf{A}$ ).

**11.1 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  $\mathbf{A}$ .

**11.2 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 dependent on the rows of  $\mathbf{X}$ .

**11.3 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:

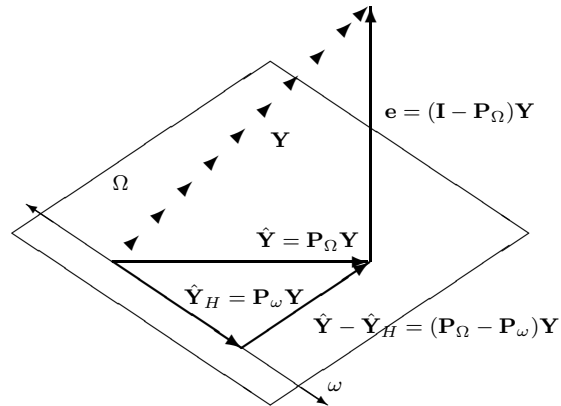
- $H : (1, 1, 0, 0)\boldsymbol{\beta} = \mu + \alpha_1 = 0$
- $H : (1, 0, 1, 0)\boldsymbol{\beta} = \mu + \alpha_2 = 0$
- $H : (1, 0, 0, 1)\boldsymbol{\beta} = \mu + \alpha_3 = 0$
- $H : (0, 1, -1, 0)\boldsymbol{\beta} = \alpha_1 - \alpha_2 = 0$
- $H : \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1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

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

$$H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0} \text{ (and } \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}) \Leftrightarrow H : \mathbf{M}\boldsymbol{\mu} = \mathbf{0} \text{ (and } \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}) \Leftrightarrow H : \boldsymbol{\mu} \in \mathcal{R}(\mathbf{X}) \cap \mathcal{N}(\mathbf{M}) \equiv \omega,$$

where  $\mathcal{N}(\mathbf{M}) = \{\mathbf{u} : \mathbf{M}\mathbf{u} = \mathbf{0}\}$  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  $\omega$  as  $\{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \mathbf{A}\boldsymbol{\beta} = \mathbf{0}\}$  or, equivalently,  $\omega = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \mathbf{M}\boldsymbol{\mu} = \mathbf{0}\}$ .



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

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

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

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

Hence

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

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

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

**11.5 Theorem:** If  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  is a testable hypothesis, then  $RSS_H - RSS = (\mathbf{A}\hat{\boldsymbol{\beta}})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}})$ .

**11.6 Theorem:** Let  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  be a testable hypothesis.

- (a)  $\text{cov}(\mathbf{A}\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}'$ .
- (b) If  $\text{rank}(\mathbf{A}) = q$ , then  $(RSS_H - RSS)/\sigma^2 = (\mathbf{A}\hat{\boldsymbol{\beta}})'[\text{cov}(\mathbf{A}\hat{\boldsymbol{\beta}})]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}})$ .
- (c)  $E[RSS_H - RSS] = \sigma^2 q + (\mathbf{A}\boldsymbol{\beta})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1}(\mathbf{A}\boldsymbol{\beta})$ .
- (d) If  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  is true and  $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ , then  $(RSS_H - RSS)/\sigma^2 \sim \chi_q^2$ .

When  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  is true,  $E[RSS_H - RSS] = \sigma^2 q$ . Therefore, we form a test statistic by calculating

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

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

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

is defined as the  $F$  distribution with  $d_1$  numerator degrees of freedom and  $d_2$  denominator degrees of freedom and is denoted  $F_{d_1, d_2}$ .

**11.8 Theorem:** 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.

**11.9 Note:** If  $\text{rank}(\mathbf{A}) = q$ , then  $\hat{\mathbf{Y}}_H = \mathbf{X}\hat{\boldsymbol{\beta}}_H$ , with

$$\hat{\boldsymbol{\beta}}_H = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1} \mathbf{A}\hat{\boldsymbol{\beta}},$$

where  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ .

**11.10 Note:** The F-test extends to  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ , for a constant  $\mathbf{c}$ . In this case, our previous results become (if  $\text{rank}(\mathbf{A}) = q$ )

$$\begin{aligned} \hat{\boldsymbol{\beta}}_H &= \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}), \\ RSS_H - RSS &= (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}). \end{aligned}$$

and  $F$  has the same distribution as before. The derivations use a solution  $\boldsymbol{\beta}_0$  to  $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{c}$  and

$$\tilde{\mathbf{Y}} \equiv \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0 = \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ .  $H$  becomes  $H : \mathbf{A}\boldsymbol{\gamma} = \mathbf{0}$ , so we can apply the previous theory to  $\tilde{\mathbf{Y}}$ .

**11.11 Example:** The t-test. Let  $U_1, \dots, U_{n_1}$  be i.i.d.  $N(\mu_1, \sigma^2)$  and  $V_1, \dots, V_{n_2}$  be i.i.d.  $N(\mu_2, \sigma^2)$ , independently of the  $U_j$ . As a linear model,

$$\begin{pmatrix} U_1 \\ \vdots \\ U_{n_1} \\ V_1 \\ \vdots \\ V_{n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_1} \\ \varepsilon_{n_1+1} \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The hypothesis  $H : \mu_1 = \mu_2$  leads to

$$F = \frac{RSS_H - RSS}{RSS/(n-2)} = (\bar{U} - \bar{V})^2 \left[ S^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{-1} = T^2,$$

where  $T = (\bar{U} - \bar{V}) / (S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$  is the two-sample  $t$  statistic.

**11.12 Example:** Multiple Linear Regression.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i.$$

The test  $H : \beta_j = 0$  ( $j \neq 0$ ) leads to

$$F = \frac{RSS_H - RSS}{RSS/(n-p)} = (\hat{\beta}_j)^2 / [\text{SE}(\hat{\beta}_j)]^2 = T^2,$$

where  $T = \hat{\beta}_j / \text{SE}(\hat{\beta}_j)$  is the usual  $t$  statistic for testing the significance of coefficients in a multiple regression model.

**11.13 Example:** Simple Linear Regression.

$$Y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i.$$

Then

$$\hat{\beta}_1 = \frac{\sum_i x_i Y_i - \sum_i x_i \sum_i Y_i / n}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_i (x_i - \bar{x})^2}$$

and

$$\text{var}(\hat{\beta}_1) = \sigma^2 / \sum_i (x_i - \bar{x})^2.$$

From the previous example, the  $F$  statistic for testing  $H : \beta_1 = 0$  is

$$F = \frac{\hat{\beta}_1^2}{S^2 / \sum_i (x_i - \bar{x})^2}.$$

It can be shown that

$$RSS = (1 - r^2) \sum_i (Y_i - \bar{Y})^2 = (1 - r^2) RSS_H,$$

where  $r$  is the sample correlation coefficient:

$$r = \frac{\sum_i (x_i - \bar{x})(Y_i - \bar{Y})}{[\sum_i (x_i - \bar{x})^2 \sum_i (Y_i - \bar{Y})^2]^{1/2}}.$$

This means that  $r^2 = (RSS_H - RSS)/RSS_H$  is the proportion of variance ( $RSS$ ) explained by the regression relationship. We will later generalize this to the sample multiple correlation coefficient ( $R^2$ ).

## 11.2 Power of the $F$ -Test:

Consider the model  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ,  $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$ , with  $\text{rank}(\mathbf{X}_{n \times p}) = r$ . Then the  $F$  statistic for testing  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  is

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

where  $\text{rank}(\mathbf{A}_{q \times p}) = q$ . Our goal is to calculate

$$\text{Power} = P(F > F_{q, n-r}^\alpha | H \text{ not true}).$$

**11.14 Definition:** Let  $X_1$  and  $X_2$  be independent random variables with  $X_1 \sim \chi_{d_1}^2(\lambda)$  and  $X_2 \sim \chi_{d_2}^2$ . Then the distribution of the ratio

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

is defined as the non-central  $F$  distribution with  $d_1$  numerator degrees of freedom,  $d_2$  denominator degrees of freedom, and non-centrality parameter  $\lambda$ , and is denoted  $F_{d_1, d_2}(\lambda)$ .

**11.15 Theorem:** The  $F$  statistic for testing  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$  has the non-central  $F$  distribution  $F \sim F_{q, n-r}(\lambda)$ , where  $\lambda = \boldsymbol{\mu}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\boldsymbol{\mu}/2\sigma^2$ .

**11.16 Note:** When calculating the non-centrality parameter  $\lambda$ , we can use the following representations:

$$\begin{aligned} \sigma^2 2\lambda &= \boldsymbol{\mu}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\boldsymbol{\mu} \\ &= \mathbf{Y}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y} |_{\mathbf{Y}=\boldsymbol{\mu}} \\ &= (RSS_H - RSS) |_{\mathbf{Y}=\boldsymbol{\mu}} \\ &= (\mathbf{A}\hat{\boldsymbol{\beta}})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}}) |_{\mathbf{Y}=\boldsymbol{\mu}} \\ &= (\mathbf{A}\boldsymbol{\beta})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\boldsymbol{\beta}). \end{aligned}$$

So we just have to substitute the true mean  $\boldsymbol{\mu}$  under the alternative hypothesis or the true parameter  $\mathbf{A}\boldsymbol{\beta}$  into appropriate formulas for  $RSS_H - RSS$ .

### 11.3 The Overall $F$ -Test:

Assume the linear model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i,$$

with full rank design matrix ( $\text{rank}(\mathbf{X}) = p$ ). Note that we are assuming the model contains an intercept. Suppose we want to test whether the overall model is significant, i.e.  $H : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ . This can be written as

$$H : \mathbf{A}\boldsymbol{\beta} = (\mathbf{0}, \mathbf{I}_{(p-1) \times (p-1)})\boldsymbol{\beta} = \mathbf{0}.$$

The  $F$  test for  $H$  yields

$$F = \frac{(RSS_H - RSS)/(p-1)}{RSS/(n-p)} \sim F_{p-1, n-p}, \quad \text{if } H \text{ is true.}$$

This is called the overall  $F$ -test statistic for the linear model. It is useful as a preliminary test of the significance of the model prior to performing model selection to determine which variables in the model are important.

### 11.4 The Multiple Correlation Coefficient:

The sample multiple correlation coefficient is defined as the correlation between the observations  $Y_i$  and the fitted values  $\hat{Y}_i$  from the regression model:

$$R \equiv \text{corr}(Y_i, \hat{Y}_i) = \frac{\sum_i (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\left[ \sum_i (Y_i - \bar{Y})^2 \sum_i (\hat{Y}_i - \bar{\hat{Y}})^2 \right]^{1/2}}.$$

**11.17 Theorem:** ANOVA decomposition:

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (Y_i - \hat{Y}_i)^2 + \sum_i (\hat{Y}_i - \bar{Y})^2 \quad \text{i.e.} \quad \text{Total SS} = \text{Residual SS} + \text{Regression SS}$$

**11.18 Theorem:**  $R^2$  as coefficient of determination:

$$R^2 = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2} = \frac{\text{REG:SS}}{\text{TOTAL:SS}},$$

or equivalently,

$$1 - R^2 = \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2} = \frac{\text{RSS}}{\text{TOTAL:SS}}.$$

**11.19 Note:**  $R^2$  is the proportion of variance in the  $Y_i$  explained by the regression model.  $R^2$  is a generalization of  $r^2$  for simple linear regression. It indicates how closely the estimated linear model fits the data. If  $R^2 = 1$  (the maximum value) then  $Y_i = \hat{Y}_i$ , and the model is a perfect fit.

**11.20 Theorem:** The F-test of a hypothesis of the form  $H : (\mathbf{0}, \mathbf{A}_1)\boldsymbol{\beta} = \mathbf{0}$  (i. e. the test does not involve the intercept  $\beta_0$ ) is a test for a significant reduction in  $R^2$ :

$$F = \frac{(R^2 - R_H^2)(n - p)}{(1 - R^2)q},$$

where  $R^2$  and  $R_H^2$  are the sample multiple correlation coefficients for the full model and the reduced model, respectively.

**11.21 Note:** This shows that  $R^2$  cannot increase when deleting a variable in the model (other than the intercept).

### 11.5 A Canonical Form for $H$ :

There are two ways to calculate the statistic  $F = \frac{(RSS_H - RSS)/q}{RSS/(n-r)}$  for a testable hypothesis  $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ .

1. Fit the full model and calculate  $RSS_H - RSS = (\mathbf{A}\hat{\boldsymbol{\beta}})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}})$ .
2. Fit the full model and calculate  $RSS$ . Then fit the reduced model and calculate  $RSS_H$ .

The reduced model is  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , with  $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ . To fit this model using a "canned" computer package, we need to represent it as

$$\mathbf{Y} = \mathbf{X}_H\boldsymbol{\beta}_H + \boldsymbol{\varepsilon}_H.$$

This is called a canonical form for  $H$ . Assume  $\text{rank}(\mathbf{A}) = q$ . Then reorder the components of  $\boldsymbol{\beta}$  and columns of  $\mathbf{A}$  so that  $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$  where  $\mathbf{A}_2$  consists of  $q$  linearly independent columns from  $\mathbf{A}$  ( $\mathbf{A}_2$  is invertible). Hence

$$H : (\mathbf{A}_1, \mathbf{A}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \mathbf{A}_1\boldsymbol{\beta}_1 + \mathbf{A}_2\boldsymbol{\beta}_2 = \mathbf{0},$$

and

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{X}_1, \mathbf{X}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}.$$

This leads to  $\mathbf{X}_H = (\mathbf{X}_1 - \mathbf{X}_2\mathbf{A}_2^{-1}\mathbf{A}_1)$  and  $\boldsymbol{\beta}_H = \boldsymbol{\beta}_1$ .

**11.22 Example:** Analysis of variance  $Y_{ij} = \alpha_i + \varepsilon_{ij}$ . In block matrix form the model is  $[\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})]$ .

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \cdots & \mathbf{1}_{n_p} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}$$

Test  $H : \alpha_1 = \alpha_2 = \cdots = \alpha_p$ , i.e.

$$\begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

A canonical form for  $H$  is  $\mathbf{Y} = \mathbf{1}\alpha_1 + \boldsymbol{\varepsilon}$ .

## 11.6 The $F$ Test for Goodness of Fit:

How can we assess if a linear model  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  is appropriate? Do the predictors adequately describe the mean of  $\mathbf{Y}$  or are there important predictors excluded? This is quite different from the overall  $F$  test which tests if the predictors are related to the response. We can test model adequacy if there are replicates, i.e. independent observations with the same values of the predictors (and so the same mean). Suppose, for  $i = 1, \dots, n$ , we have replicates  $Y_{i1}, \dots, Y_{iR_i}$  corresponding to the values  $x_{i1}, \dots, x_{i,p-1}$  of the predictors. The full model is

$$Y_{ir} = \mu_i + \varepsilon_{ir}$$

where the  $\mu_i$  are any constants. We wish to test whether they have the form

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1}.$$

If  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ , we want to test the hypothesis

$$H : \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}.$$

We now apply the general  $F$  test to  $H$ . The RSS under the full model is

$$RSS = \sum_{i=1}^n \sum_{r=1}^{R_i} (Y_{ir} - \bar{Y}_i)^2,$$

and for the reduced model

$$RSS_H = \sum_{i=1}^n \sum_{r=1}^{R_i} (Y_{ir} - \hat{\beta}_{0H} - \hat{\beta}_{1H} x_{i1} - \dots - \hat{\beta}_{p-1,H} x_{i,p-1})^2.$$

It can be shown that in the case  $R_i = R$  the estimates under the reduced model are

$$\hat{\boldsymbol{\beta}}_H = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z},$$

where  $Z_i = \bar{Y}_i = \sum_{r=1}^R Y_{ir}/R$ . The  $F$  statistic is

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

where  $N = \sum_{i=1}^n R_i$ . This test is also called the lack-of-fit test.