trace I_n = trace $\sum A_i = \sum$ trace A_i

LEMMA

(3)$\rightarrow$(1) $n = \text{trace}\, I_n = \sum \text{trace}\, A_i \overset{12}{=} \sum \text{rank}\, A_i = \sum r_i$

(1)$\rightarrow$(2) Let $V_i = \{A_i\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} = \text{span}(A_i)$. Then $\dim V_i = r_i$. Let B_i be a basis for V_i and let $B = \bigcup_i B_i$. Since $\mathbf{x} = I\mathbf{x} = \sum A_i\mathbf{x}$, $\forall \mathbf{x} \in \mathbb{R}^n$, B spans $\mathbb{R}^n$ and since B has at most $\sum r_i = n$ elements, B must form a basis of $\mathbb{R}^n$. Hence, any $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as $\sum \mathbf{u}_i$ where $\mathbf{u}_i \in V_i$. Let $\mathbf{x}$ be a column of A_j. Then $\underbrace{\mathbf{x}}_{\in V_j} + \sum_{i \neq j} \mathbf{0} = \sum A_i\mathbf{x}$. By uniqueness, $A_i\mathbf{x} = \mathbf{0}$ for all $i \neq j$. $A_i x = 0$

$A_i x = 0$ $\forall x$ column of A_J

$\Rightarrow A_i A_J = 0$ $\forall i \neq J$

Theorem 9 (The Fisher-Cochran Theorem)
If $A_1, \dots, A_k$ are $n \times n$ projection matrices such that $\sum_{i=1}^n A_i = I_n$, and if $\mathbf{Z} \sim N(\boldsymbol{\mu}, I_n)$ then $\mathbf{Z}^T A_1 \mathbf{Z}, \dots, \mathbf{Z}^T A_k \mathbf{Z}$ are independent and

$$\mathbf{Z}^T A_i \mathbf{Z} \sim \chi^2_{r_i}(\delta_i), \quad \text{where } r_i = \text{rank}\, A_i \text{ and } \delta_i^2 = \boldsymbol{\mu}^T A_i \boldsymbol{\mu}.$$

BY LEMMA 13

Proof By Lemma 20, $A_iA_j = 0$ for all $i \neq j$. Hence, $\mathbf{Z}^T A_1 \mathbf{Z}, \dots, \mathbf{Z}^T A_k \mathbf{Z}$ are independent.

The rest of the theorem is a consequence of Lemma 18.

10.2 The Linear Model with Normal Theory Assumptions

In this section we will consider the linear model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\text{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ with (NTA).

Recall that the (NTA) are $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$. In particular, this implies $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$. The joint probability density function of $\mathbf{Y}$ is thus

$$f(\mathbf{y}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta})\right)$$

Estimation using the maximum likelihood approach:

- The log-likelihood of the data is

$$L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\underbrace{(\mathbf{Y} - X\boldsymbol{\beta})^T(\mathbf{Y} - X\boldsymbol{\beta})}_{=S(\boldsymbol{\beta})}$$

$B = \bigcup_{i=1}^{n} e_i \qquad e_i = (0,0,\dots,0,\underset{\substack{\uparrow \\ i\text{-th}}}{1},0,\dots,0)$

$B_1 = e_1 \cup e_2 \cup e_3$

$B_2 = e_4 \cup e_5$

$x \in \mathbb{R}^n \qquad x = (3,4,9,7,8,\dots,10)$

$u_i \in V_i$ $\quad u_i$ CAN BE WRITTEN UNIQUELY AS A LINEAR COMBINATION OF THE ELEMENTS OF B_i

$u_1 = (3,4,9,0,0,\dots,0) \qquad u_2 = (0,0,0,7,8,0,\dots,0)$

$$\left[\begin{array}{l} y = (1,4,5,0,\dots,0) \\ \tilde{u}_2 = 0 = \tilde{u}_3 = \dots = \tilde{u}_k \end{array}\right] \Rightarrow y \in V_1$$

Remark Suppose we construct a tests with the above pivotal quantity for $\mathbf{c}^T\beta$. It turns out that the test statistic has a non-central t-distribution under the alternative hypothesis.

→ EXERCISE

10.4 The F-Test

$c^T\beta \in \mathbb{R}$

In the previous section, we derived pivotal quantities for one-dimensional parameters (σ^2 or linear combinations $\mathbf{c}^T\beta$ of the components of β such as, for some i, $\mathbf{e}_i^T\beta = \beta_i$). If we are interested in how more than one component of the parameter behaves, e.g. if the null-hypotheses $\beta_2 = \beta_3 = 0$ is of interest then we would have to do more than one test (and this would result in similar problems as the "joint confidence intervals" mentioned earlier and a correction such as the Bonferroni correction would be necessary). This section presents a method to test more complicated hypotheses about β.

Example 60
Suppose we have a linear model with $p = 3$ and design matrix

$$X = \begin{pmatrix} 1 & a_1 & b_1 \\ \vdots & \vdots & \vdots \\ 1 & a_n & b_n \end{pmatrix}$$

$Y_i = \beta_1 + \beta_2 a_1 + \beta_3 b_1 + \varepsilon_i$

Suppose we are interested in testing the hypotheses

$Y_i = \beta_1 + \varepsilon_i$

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{against} \quad H_1 : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

Under H_0, we can write the linear model as

$$\mathrm{E}\, Y = X_0\beta_1, \quad \text{where } X_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Thus we can rewrite the hypotheses as

$$H_0 : \mathrm{E}\,\mathbf{Y} \in \operatorname{span}(X_0) \quad \text{against} \quad H_1 : \mathrm{E}\,\mathbf{Y} \notin \operatorname{span}(X_0)$$

WHERE X_0 IS THE DESIGN MATRIX OF THE RESTRICTED MODEL

In general, suppose we want to test whether a sub-model of a linear model $\mathrm{E}\,\mathbf{Y} = X\beta$ is true, i.e. we want to test

$$H_0 \;:\; \mathrm{E}\,\mathbf{Y} \in \operatorname{span}(X_0) \text{ against } H_1 \;:\; \mathrm{E}\,\mathbf{Y} \notin \operatorname{span}(X_0)$$

for some matrix X_0 with $\boxed{\operatorname{span}(X_0) \subset \operatorname{span}(X)}$. In other words, the null hypothesis says that the sub-model

$$\mathrm{E}(\mathbf{Y}) = X_0\beta_0$$

is true.

$H_0 \to Y_i = \beta_1 + \beta_2 a_i + \varepsilon_i$ VS

$H_1 \to Y_i = \beta_1 + \beta_2 a_i + \beta_3 b_i + \varepsilon_i$

VS

Example 61
Continuing the previous example, one may also be interested in $X_0 = \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_n \end{pmatrix}$ [equivalent to $\beta_3 = 0$] or $X_0 = \begin{pmatrix} 1 & a_1 - b_1 \\ \vdots & \vdots \\ 1 & a_n - b_n \end{pmatrix}$ [equivalent to $\beta_3 = -\beta_2$].

$H_0: \; Y_i = \beta_1 + (a_i - b_i)\beta_2 + \varepsilon_i \; = \beta_1 + a_i\beta_2 - b_i\beta_2 + \varepsilon_i$

$\beta_3 = -\beta_2$

Let

- RSS= the residual sum of squares in the full model $\mathrm{E}\,\mathbf{Y} = X\beta$
- RSS_0= the residual sum of squares in the sub-model $\mathrm{E}\,\mathbf{Y} = X_0\tilde{\beta}$

Lemma 23
Under $H_0 \;:\; \mathrm{E}\,\mathbf{Y} \in \operatorname{span}(X_0)$,

$$F = \frac{\mathrm{RSS}_0 - \mathrm{RSS}}{\mathrm{RSS}} \cdot \frac{n-r}{r-s} \sim F_{r-s,n-r}$$

where $r = \operatorname{rank} X$, $s = \operatorname{rank} X_0$.

Let P be the projection matrix onto span X, and let $Q = I - P$ the projection matrix onto $(\operatorname{span} X)^{\perp}$.
Likewise, let P_0 be the projection matrix onto span X_0, and let $Q_0 = I - P_0$ the projection matrix onto $(\operatorname{span} X_0)^{\perp}$. Then, as we derived in the previous chapter,

$$\mathrm{RSS} = \mathbf{Y}^T Q \mathbf{Y}, \quad \mathrm{RSS}_0 = \mathbf{Y}^T Q_0 \mathbf{Y}.$$

Using this gives

$Q_0 - Q = I - P_0 - (I - P) =$
$= P - P_0$

$$F = \frac{\mathbf{Y}^T Q_0 \mathbf{Y} - \mathbf{Y}^T Q \mathbf{Y}}{\mathbf{Y}^T Q \mathbf{Y}} \cdot \frac{n-r}{r-s}$$
$$= \frac{\mathbf{Y}^T (P - P_0) \mathbf{Y}/\sigma^2}{\mathbf{Y}^T (I - P) \mathbf{Y}/\sigma^2} \cdot \frac{n-r}{r-s}$$

To show: numerator$\sim \chi^2_{r-s}$, denominator$\sim \chi^2_{n-r}$, numerator and denominator are independent.

Proof We will use the Fisher-Cochran theorem. Let $\mathbf{Z} = \mathbf{Y}/\sigma$, $A_1 = I - P$, $A_2 = P - P_0$, $A_3 = P_0$.

Clearly, $A_1 + A_2 + A_3 = I$. We already know that A_1 and A_3 are projection matrices.

To show: $A_2 = P - P_0$ is a projection matrix. $P - P_0$ is symmetric as P_0 and P are both symmetric. Furthermore,

$$(P - P_0)^2 = P^2 + P_0^2 - PP_0 - P_0P \quad = P - P_0$$

Every column $\mathbf{y}$ of P_0 is an element of span(X_0) and thus an element of span(X). Thus, $P\mathbf{y} = \mathbf{y}$.
Hence,

$$PP_0 = P_0.$$

P_0 IS SYMMETRIC

This also implies $P_0 P = (P^T P_0^T)^T = (PP_0)^T = P_0^T = P_0$.

SYMMETRIC OF P AND P_0

Thus,

$$(P - P_0)^2 = P + P_0 - P_0 - P_0 = P - P_0$$

The Fisher-Cochran theorem now implies

- $\mathbf{Z}^T(P - P_0)\mathbf{Z}$ and $\mathbf{Z}^T(I - P)\mathbf{Z}$ are independent,
- $\mathbf{Z}^T(P - P_0)\mathbf{Z} \sim \chi^2_{\text{rank}(P-P_0)}(\mathrm{E}\,\mathbf{Z}^T(P - P_0)\,\mathrm{E}\,\mathbf{Z})$,
- $\mathbf{Z}^T(I - P)\mathbf{Z} \sim \chi^2_{\text{rank}(I-P)}(\mathrm{E}\,\mathbf{Z}^T(I - P)\,\mathrm{E}\,\mathbf{Z})$.

Next, we show that the non-centrality parameters are 0.

Under H_0, we know $\mathrm{E}\,\mathbf{Z} = \frac{1}{\sigma}\mathrm{E}\,\mathbf{Y} \in \text{span}(X_0) \subset \text{span}(X)$ Thus,

$$(P - P_0)\,\mathrm{E}\,\mathbf{Z} = \underbrace{P\,\mathrm{E}\,\mathbf{Z}}_{=\mathrm{E}\,\mathbf{Z}} - \underbrace{P_0\,\mathrm{E}\,\mathbf{Z}}_{=\mathrm{E}\,\mathbf{Z}} = \mathbf{0}.$$

HERE IS THE ONLY PLACE WHERE WE ~~WE USE~~ USE THAT WE ARE UNDER H_0

Hence, $\mathrm{E}\,\mathbf{Z}^T(P-P_0)\,\mathrm{E}\,\mathbf{Z}=0$. Furthermore,

$$E\mathbf{Z}^T(I-P)\,\mathrm{E}\,\mathbf{Z} = E\mathbf{Z}^T(\mathrm{E}\,\mathbf{Z} - \underbrace{P\,\mathrm{E}\,\mathbf{Z}}_{=\mathrm{E}\,\mathbf{Z}}) = 0.$$

Concerning the degrees of freedom:

- By Lemma 20, $n = \sum n_i$

  $$n = \operatorname{rank}(P) + \operatorname{rank}(I-P) = \operatorname{rank} X + \operatorname{rank}(I-P) = r + \operatorname{rank}(I-P)$$

  Thus, $\operatorname{rank}(I-P) = n-r$.

- Using Lemma 20 again,

  $$\underbrace{\operatorname{rank}(P_0)}_{=\operatorname{rank}(X_0)=s} + \operatorname{rank}(P-P_0) + \underbrace{\operatorname{rank}(I-P)}_{n-r} = n$$

  Thus, $\operatorname{rank}(P-P_0) = r-s$.

To summarise, we have shown

$$\mathbf{Z}^T(P-P_0)\mathbf{Z} \sim \chi^2_{r-s},\ \mathbf{Z}^T(I-P)\mathbf{Z} \sim \chi^2_{n-r} \text{ independently.}$$

Thus, by definition, $F \sim F_{r-s,n-r}$.

If H_0 is not true then the proof is still valid, except for the non-centrality parameter of $\mathbf{Z}^T(P-P_0)\mathbf{Z}$. Now,

$EY = X\beta$

$$\mathrm{E}\,\mathbf{Z}^T(P-P_0)\,\mathrm{E}\,\mathbf{Z} = \frac{1}{\sigma^2}\beta^T X^T(P-P_0)X\beta.$$

Thus, without assuming H_0, we get

$$F \sim F_{r-s,n-r}(\delta), \text{ where } \delta^2 = \frac{1}{\sigma^2}(X\beta)^T(P-P_0)X\beta.$$

This implies that F will take on larger values if H_0 is not true.

Thus it is advisable to reject for large values of F. In particular, if we want a test to the level $\alpha > 0$, we reject if

$$F > c,$$

where c is such that $\mathrm{P}(X \geq c) = \alpha$ for $X \sim F_{r-s,n-r}$.

UNDER H_0