Proof All nonzero eigenvalues of A are equal to 1.

By Lemma 16, $\exists\ L \in \mathbb{R}^{n\times r}$ such that $A = LL^T$ and $L^T L = I_r$. Let $\mathbf{V} = L^T\mathbf{Z}$. Then $\mathbf{V} \sim N(L^T\boldsymbol{\mu}, \underbrace{I}_{=L^T L})$ and

$$\mathbf{Z}^T A\mathbf{Z} = \mathbf{Z}^T LL^T\mathbf{Z} = \mathbf{V}^T\mathbf{V} \sim \chi_r^2(\delta),$$

where $\delta^2 = \left(L^T\boldsymbol{\mu}\right)^T L^T\boldsymbol{\mu} = \boldsymbol{\mu}^T \underbrace{LL^T}_{=A}\boldsymbol{\mu} = \boldsymbol{\mu}^T A\boldsymbol{\mu}$.

> **Lemma 19**
> If $\mathbf{Z} \sim N(\boldsymbol{\mu}, I_n)$ and $A_1,\ A_2 \in \mathbb{R}^{n\times n}$ are projection matrices and $A_1A_2 = 0$ then $\mathbf{Z}^T A_1\mathbf{Z}$ and $\mathbf{Z}^T A_2\mathbf{Z}$ are independent.

Proof $\mathrm{cov}(A_1\mathbf{Z}, A_2\mathbf{Z}) = A_1 \underbrace{\mathrm{cov}(\mathbf{Z},\mathbf{Z})}_{=I} A_2^T = A_1A_2^T = 0$.

Because $A_1\mathbf{Z}$ and $A_2\mathbf{Z}$ are jointly normally distributed this shows that they are independent.

As $\mathbf{Z}^T A_i\mathbf{Z} = (A_i\mathbf{Z})^T(A_i\mathbf{Z})$ for $i = 1, 2$ (symm+ idempotent) this implies that $\mathbf{Z}^T A_1\mathbf{Z}$ and $\mathbf{Z}^T A_2\mathbf{Z}$ are independent.

This result extends to $\mathbf{Z}^T A_1\mathbf{Z}, \dots, \mathbf{Z}^T A_k\mathbf{Z}$, where $A_iA_j = 0$ $(i \neq j)$.

> **Lemma 20**
> If $A_1, \dots, A_k$ are symmetric $n \times n$ matrices such that $\sum A_i = I_n$ and if rank $A_i = r_i$ then the following are equivalent:
>
> 1. $\sum r_i = n$
> 2. $A_iA_j = 0$ for all $i \neq j$
> 3. A_i is idempotent for all $i = 1, \dots, k$.

MULTIPLY BY A_J

Proof (2)→(3) $\forall j$: $A_1 + \cdots + A_k = I_n \implies \underbrace{A_1A_j}_{=0} + \cdots + \underbrace{A_kA_j}_{=0} = A_j \implies A_j^2 = A_j$.

$A_JA_J = A_J^2$

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

LEMMA

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

(1)→(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

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

- maximising L with respect to β (for fixed σ^2) is equivalent to minimising $S(\beta) = (\mathbf{Y} - X\beta)^T(\mathbf{Y} - X\beta)$, i.e. maximum likelihood is equivalent to least squares for estimating β.

- The maximum likelihood estimator for σ^2 is RSS $/n$ (look at first/second derivative of $L(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}$ RSS wrt σ^2).

10.3 Confidence Intervals, Tests for one-dimensional quantities

The following gives a pivotal quantity for σ^2 which can be used to construct CIs or tests.

Lemma 21 (Distribution of RSS)
Assume (NTA). Then

$$\frac{\text{RSS}}{\sigma^2} \sim \chi^2_{n-r}$$

where $r = \operatorname{rank} X$.

$Y \sim N(X\beta, \sigma^2 I)$

Proof Let P denote the projection matrix onto span(X). Then

$$\text{RSS} = \mathbf{e}^T\mathbf{e} = (\underbrace{(I-P)}_{=:Q}\mathbf{Y})^T(I-P)\mathbf{Y} = \mathbf{Y}^T \underbrace{Q^TQ}_{=QQ=Q} \mathbf{Y} = \mathbf{Y}^T Q\mathbf{Y}$$

and Q is the projection onto the space orthogonal to the columns of X. Hence,

$$\frac{\text{RSS}}{\sigma^2} = \frac{\mathbf{Y}}{\sigma}^T Q\frac{\mathbf{Y}}{\sigma} = \mathbf{Z}^T Q\mathbf{Z}$$

HAT MATRIX

where $\mathbf{Z} = \mathbf{Y}/\sigma \sim N(X\beta/\sigma, I)$ and Q is a projection matrix. Furthermore, $Q + P = I$ and Q and P are projection matrices. Thus, by Lemma 20, $\operatorname{rank} Q + \underbrace{\operatorname{rank} P}_{=r} = n$, implying $\operatorname{rank} Q = n - r$.

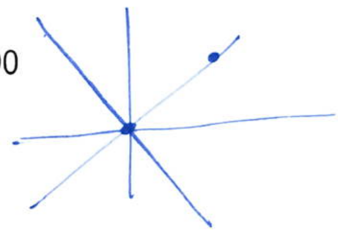
Thus, by Lemma 18,

$$\frac{\text{RSS}}{\sigma^2} \sim \chi^2_{n-r}(0)$$

since $(X\beta/\sigma)^T \underbrace{QX}_{=0} \beta/\sigma = 0$.

$\delta^2 = \mu^T A_i \mu$

$\frac{(X\beta)^T}{\sigma} Q \frac{X\beta}{\sigma}$



Often, we will be interested in parts of the parameter vector β, e.g. one of its components β_i. We want to construct tests, or construct confidence intervals.

The following lemma provides a flexible way of doing this; it gives a pivotal quantity for $\mathbf{c}^T\beta$ for some (known) $\mathbf{c} \in \mathbb{R}^p$. We have worked with $\mathbf{c}^T\beta$ already in the Gauss-Markov Theorem.

Lemma 22
Assume (FR), (NTA) in a linear model. Let $\mathbf{c} \in \mathbb{R}^p$. Then

$$\frac{\mathbf{c}^T\hat{\beta} - \mathbf{c}^T\beta}{\sqrt{\mathbf{c}^T(X^TX)^{-1}\mathbf{c}\frac{\text{RSS}}{n-p}}} \sim t_{n-p}$$

$p = \operatorname{rank}(X)$
(SOMETIMES WE USE r INSTEAD OF p NOTATIONWISE)

Proof Since $\mathbf{c}^T\hat{\beta} = \mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y}$ and $\mathbf{Y} \sim N(X\beta, \sigma^2 I)$,

$$\mathrm{E}\,\mathbf{c}^T\hat{\beta} = \mathbf{c}^T\beta$$

$$\begin{aligned}\operatorname{Var}(\mathbf{c}^T\hat{\beta}) =& \operatorname{Var}(\mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y}) = \mathbf{c}^T(X^TX)^{-1}X^T\underbrace{\operatorname{cov}(\mathbf{Y})}_{=\sigma^2 I}X(X^TX)^{-1}\mathbf{c}\\ =& \mathbf{c}^T(X^TX)^{-1}\mathbf{c}\sigma^2\end{aligned}$$

and thus $\mathbf{c}^T\hat{\beta} \sim N(\mathbf{c}^T\beta, \mathbf{c}^T(X^TX)^{-1}\mathbf{c}\sigma^2)$. Hence,

$$\frac{\mathbf{c}^T\hat{\beta} - \mathbf{c}^T\beta}{\sqrt{\mathbf{c}^T(X^TX)^{-1}\mathbf{c}\sigma^2}} \sim N(0,1).$$

$$\frac{X}{\sqrt{\frac{W}{n}}} \qquad W \to \frac{RSS}{\sigma^2} \sim \chi^2_{n-p} \qquad n \to n-p$$

We already know $\text{RSS}/\sigma^2 \sim \chi^2_{n-p}$.

Thus the lemma is a consequence of the definition of the t-distribution once we have shown that $\hat{\beta} = (X^TX)^{-1}X^T\mathbf{Y}$ and $\text{RSS} = \mathbf{Y}^TQ\mathbf{Y}$ are independent. The latter is a consequence of Lemma 17 (using $\mathbf{Z} = \mathbf{Y}/\sigma$), since $\underbrace{(X^TX)^{-1}X^T}_{B}\underbrace{Q}_{A} = (X^TX)^{-1}(\underbrace{QX}_{=0})^T = 0$.

Example 59
Data Set: Tooth Growth (see File in Additional Material)

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

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

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

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

$Y_i = \beta_1 + \varepsilon_i$

Suppose we are interested in testing the hypotheses

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

Under H_0, we can write the linear model as

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

Thus we can rewrite the hypotheses as

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