Lecture 15. Hypothesis testing in the linear model

Preliminary lemma

Lemma 15.1

Suppose $\mathbf{Z} \sim N_n(\mathbf{0}, \sigma^2 I_n)$ and A_1 and A_2 and symmetric, idempotent $n \times n$ matrices with $A_1 A_2 = 0$. Then $\mathbf{Z}^T A_1 \mathbf{Z}$ and $\mathbf{Z}^T A_2 \mathbf{Z}$ are independent.

Proof:

- Let $\mathbf{W}_i = A_i \mathbf{Z}, i = 1, 2$ and $\underset{2n\times 1}{\mathbf{W}} = \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{pmatrix} = A\mathbf{Z}$, where $\underset{2n\times n}{A} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$.
- By Proposition 11.1(i), $\mathbf{W} \sim \mathrm{N}_{2n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma^2 \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \right)$ check.
- So $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent, which implies $\mathbf{W}_1{}^T \mathbf{W}_1 = \mathbf{Z}^T A_1 \mathbf{Z}$ and $\mathbf{W}_2{}^T \mathbf{W}_2 = \mathbf{Z}^T A_2 \mathbf{Z}$ are independent. $\square$.

Hypothesis testing

- Suppose $\underset{n\times p}{X} = (\underset{n\times p_0}{X_0} \;\; \underset{n\times(p-p_0)}{X_1})$ and $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$, where $\mathrm{rank}(X) = p, \mathrm{rank}(X_0) = p_0$.
- We want to test $H_0 : \boldsymbol{\beta}_1 = 0$ against $H_1 : \boldsymbol{\beta}_1 \neq 0$.
- Under H_0, $\mathbf{Y} = X_0 \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$.
- Under H_0, MLEs of $\boldsymbol{\beta}_0$ and σ^2 are

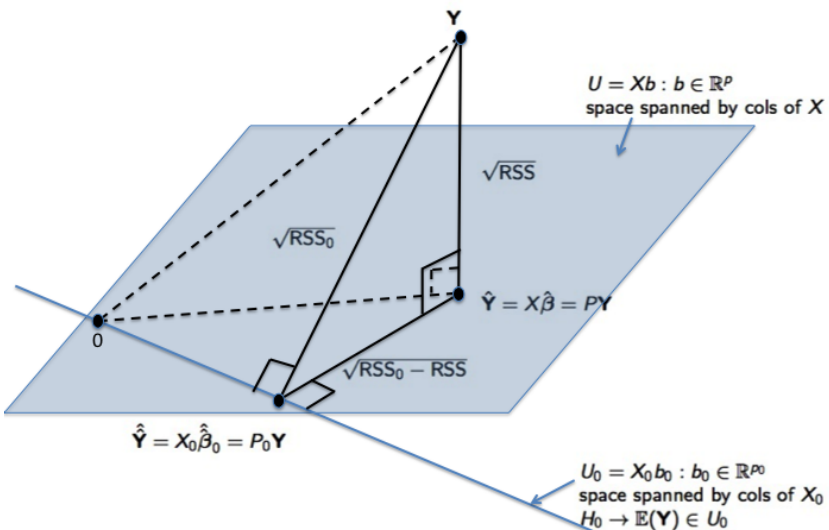
$$\begin{aligned} \hat{\hat{\boldsymbol{\beta}}}_0 &= (X_0{}^T X_0)^{-1} X_0{}^T \mathbf{Y} \\ \hat{\hat{\sigma}}^2 &= \frac{\mathrm{RSS}_0}{n} = \frac{1}{n}(\mathbf{Y} - X_0 \hat{\hat{\boldsymbol{\beta}}}_0)^T (\mathbf{Y} - X_0 \hat{\hat{\boldsymbol{\beta}}}_0) \end{aligned}$$

and these are independent, by Theorem 13.3.
- So fitted values under H_0 are

$$\hat{\hat{\mathbf{Y}}} = X_0 (X_0{}^T X_0)^{-1} X_0{}^T \mathbf{Y} = P_0 \mathbf{Y},$$

where $P_0 = X_0 (X_0{}^T X_0)^{-1} X_0{}^T$.

Geometric interpretation



Generalised likelihood ratio test

- The generalised likelihood ratio test of H_0 against H_1 is

$$\Lambda_{\mathbf{Y}}(H_0, H_1) = \frac{\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n \exp\left(-\frac{1}{2\hat{\sigma}^2}(\mathbf{Y}-X\hat{\beta})^T(\mathbf{Y}-X\hat{\beta})\right)}{\left(\frac{1}{\sqrt{2\pi\hat{\hat{\sigma}}^2}}\right)^n \exp\left(-\frac{1}{2\hat{\hat{\sigma}}^2}(\mathbf{Y}-X\hat{\hat{\beta}}_0)^T(\mathbf{Y}-X\hat{\hat{\beta}}_0)\right)}$$
$$= \left(\frac{\hat{\hat{\sigma}}^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{\text{RSS}_0}{\text{RSS}}\right)^{\frac{n}{2}} = \left(1+\frac{\text{RSS}_0-\text{RSS}}{\text{RSS}}\right)^{\frac{n}{2}}$$

- We reject H_0 when $2\log\Lambda$ is large, equivalently when $\frac{(\text{RSS}_0-\text{RSS})}{\text{RSS}}$ is large.
- Using the results in Lecture 8, under H_0

$$2\log\Lambda = n\log\left(1+\frac{\text{RSS}_0-\text{RSS}}{\text{RSS}}\right)$$

is approximately a $\chi^2_{p_1-p_0}$ rv.
- But we can get an exact null distribution.

Null distribution of test statistic

- We have $\text{RSS} = \mathbf{Y}^T(I_n-P)\mathbf{Y}$ (see proof of Theorem 13.3 (ii)), and so

$$\text{RSS}_0 - \text{RSS} = \mathbf{Y}^T(I_n-P_0)\mathbf{Y} - \mathbf{Y}^T(I_n-P)\mathbf{Y} = \mathbf{Y}^T(P-P_0)\mathbf{Y}.$$

- Now I_n-P and $P-P_0$ are symmetric and idempotent, and therefore $\text{rank}(I_n-P) = n-p$, and

$$\text{rank}(P-P_0) = \text{tr}(P-P_0) = \text{tr}(P) - \text{tr}(P_0) = \text{rank}(P) - \text{rank}(P_0) = p - p_0.$$

- Also

$$(I_n-P)(P-P_0) = (I_n-P)P - (I_n-P)P_0 = 0.$$

- Finally,

$$\begin{aligned}\mathbf{Y}^T(I_n-P)\mathbf{Y} &= (\mathbf{Y}-X_0\boldsymbol{\beta}_0)^T(I_n-P)(\mathbf{Y}-X_0\boldsymbol{\beta}_0) \text{ since } (I_n-P)X_0 = 0,\\ \mathbf{Y}^T(P-P_0)\mathbf{Y} &= (\mathbf{Y}-X_0\boldsymbol{\beta}_0)^T(P-P_0)(\mathbf{Y}-X_0\boldsymbol{\beta}_0) \text{ since } (P-P_0)X_0 = 0,\end{aligned}$$

- Applying Lemmas 13.2 ($\mathbf{Z}^T A_i \mathbf{Z} \sim \sigma^2\chi^2_r$) and 15.1 to $\mathbf{Z} = \mathbf{Y} - X_0\boldsymbol{\beta}_0, A_1 = I_n - P, A_2 = P - P_0$ to get that under H_0,

$$\begin{aligned}\text{RSS} = \mathbf{Y}^T(I_n-P)\mathbf{Y} &\sim \chi^2_{n-p}\\ \text{RSS}_0 - \text{RSS} = \mathbf{Y}^T(P-P_0)\mathbf{Y} &\sim \chi^2_{p-p_0}\end{aligned}$$

and these rvs are independent.
- So under H_0,

$$F = \frac{\mathbf{Y}^T(P-P_0)\mathbf{Y}/(p-p_0)}{\mathbf{Y}^T(I_n-P)\mathbf{Y}/(n-p)} = \frac{(\text{RSS}_0-\text{RSS})/(p-p_0)}{\text{RSS}/(n-p)} \sim F_{p-p_0,n-p}.$$

- Hence we reject H_0 if $F > F_{p-p_0,n-p}(\alpha)$.
- RSS_0 - RSS is the 'reduction in the sum of squares due to fitting $\boldsymbol{\beta}_1$.

Arrangement as an 'analysis of variance' table

Source of variation	degrees of freedom (df)	sum of squares	mean square	F statistic
Fitted model	$p-p_0$	RSS_0 - RSS	$\frac{(\text{RSS}_0-\text{RSS})}{(p-p_0)}$	$\frac{(\text{RSS}_0-\text{RSS})/(p-p_0)}{\text{RSS}/(n-p)}$
Residual	$n-p$	RSS	$\frac{\text{RSS}}{(n-p)}$	
	$n-p_0$	RSS_0		

The ratio $\frac{(\text{RSS}_0-\text{RSS})}{\text{RSS}_0}$ is sometimes known as the *proportion of variance explained by* $\boldsymbol{\beta}_1$, and denoted R^2.

Simple linear regression

- We assume that
$$Y_i = a' + b(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, \dots, n,$$
where $\bar{x} = \sum x_i/n$, and $\varepsilon_i, i = 1, ..., n$ are iid $N(0, \sigma^2)$.
- Suppose we want to test the hypothesis $H_0 : b = 0$, i.e. no linear relationship. From Lecture 14 we have seen how to construct a confidence interval, and so could simply see if it included 0.
- Alternatively , under H_0, the model is $Y_i \sim N(a', \sigma^2)$, and so $\hat{a}' = \overline{Y}$, and the fitted values are $\hat{Y}_i = \overline{Y}$.
- The observed RSS_0 is therefore
$$\text{RSS}_0 = \sum_i (y_i - \overline{y})^2 = S_{yy}.$$
- The fitted sum of squares is therefore
$$\text{RSS}_0 - \text{RSS} = \sum_i \left((y_i - \overline{y})^2 - (y_i - \overline{y} - \hat{b}(x_i - \bar{x}))^2 \right) = \hat{b}^2 (x_i - \bar{x})^2 = \hat{b}^2 S_{xx}.$$

Source of variation	d.f.	sum of squares	mean square	F statistic
Fitted model	1	$\text{RSS}_0 - \text{RSS} = \hat{b}^2 S_{xx}$	$\hat{b}^2 S_{xx}$	$F = \hat{b}^2 S_{xx}/\tilde{\sigma}^2$
Residual	$n-2$	$\text{RSS} = \sum_i (y_i - \hat{y})^2$	$\tilde{\sigma}^2$	
	$n-1$	$\text{RSS}_0 = \sum_i (y_i - \overline{y})^2$		

- Note that the proportion of variance explained is $\hat{b}^2 S_{xx}/S_{yy} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = r^2$, where r is Pearson's Product Moment Correlation coefficient $r = S_{xy}/\sqrt{S_{xx}S_{yy}}$.
- From lecture 14, slide 5, we see that under H_0, $\frac{\hat{b}}{\text{s.e.}(\hat{b})} \sim t_{n-2}$, where $\text{s.e.}(\hat{b}) = \tilde{\sigma}/\sqrt{S_{xx}}$.
So $\frac{\hat{b}}{\text{s.e.}(\hat{b})} = \frac{\hat{b}\sqrt{S_{xx}}}{\tilde{\sigma}} = t$.
- Checking whether $|t| > t_{n-2}(\frac{\alpha}{2})$ is precisely the same as checking whether $t^2 = F > F_{1,n-2}(\alpha)$, since a $F_{1,n-2}$ variable is t_{n-2}^2.
- Hence the same conclusion is reached, whether based on a t-distribution or the F statistic derived from an analysis-of-variance table.

Example 12.1 continued

As R code

```
> fit=lm(time~ oxy.s  )
> summary.aov(fit)

            Df Sum Sq Mean Sq F value   Pr(>F)
oxy.s        1 129690  129690   41.98 1.62e-06 ***
Residuals   22  67968    3089
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Note that the F statistic, 41.98, is -6.48^2, the square of the t statistic on Slide 5 in Lecture 14.

One way analysis of variance with equal numbers in each group

- Assume J measurements taken in each of I groups, and that
$$Y_{i,j} = \mu_i + \varepsilon_{i,j},$$
where $\varepsilon_{i,j}$ are independent $N(0, \sigma^2)$ random variables, and the μ_i's are unknown constants.
- Fitting this model gives $\text{RSS} = \sum_{i=1}^I \sum_{j=1}^J (Y_{i,j} - \hat{\mu}_i)^2 = \sum_{i=1}^I \sum_{j=1}^J (Y_{i,j} - \overline{Y}_{i.})^2$ on $n - I$ degrees of freedom.
- Suppose we want to test the hypothesis $H_0 : \mu_i = \mu$, i.e. no difference between groups.
- Under H_0, the model is $Y_{i,j} \sim N(\mu, \sigma^2)$, and so $\hat{\mu} = \overline{Y}_{..}$, and the fitted values are $\hat{Y}_{i,j} = \overline{Y}_{..}$.
- The observed RSS_0 is therefore
$$\text{RSS}_0 = \sum_i \sum_j (y_{i,j} - \overline{y}_{..})^2.$$

- The fitted sum of squares is therefore

$$\mathrm{RSS}_0 - \mathrm{RSS} = \sum_i \sum_j \left((y_{i,j} - \overline{y}_{..})^2 - (y_{i,j} - \overline{y}_{i.})^2 \right) = J \sum_i (\overline{y}_{i.} - \overline{y}_{..})^2.$$

Source of variation	d.f.	sum of squares	mean square	F statistic
Fitted model	$I-1$	$J\sum_i(\overline{y}_{i.} - \overline{y}_{..})^2$	$\frac{J\sum_i(\overline{y}_{i.} - \overline{y}_{..})^2}{(I-1)}$	$F = \frac{J\sum_i(\overline{y}_{i.} - \overline{y}_{..})^2}{(I-1)\tilde{\sigma}^2}$
Residual	$n-I$	$\sum_i\sum_j(y_{i,j} - \overline{y}_{i.})^2$	$\tilde{\sigma}^2$	
	$n-1$	$\sum_i\sum_j(y_{i,j} - \overline{y}_{..})^2$		

Example 13.1

As R code

```
> summary.aov(fit)

            Df Sum Sq Mean Sq F value Pr(>F)
x            4  507.9   127.0    1.17  0.354
Residuals   20 2170.1   108.5
```

The p-value is 0.35, and so there is no evidence for a difference between the instruments.