Lecture 15. Hypothesis testing in the linear model
15. Hypothesis testing in the linear model

15.1. Preliminary lemma

**Lemma 15.1**

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

**Proof:**

- Let \( W_i = A_i Z, i = 1, 2 \) and \( W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = A Z \), where \( A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \).
- By Proposition 11.1(i), \( W \sim N_{2n} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \right) \) check.
- So \( W_1 \) and \( W_2 \) are independent, which implies \( W_1^T W_1 = Z^T A_1 Z \) and \( W_2^T W_2 = Z^T A_2 Z \) are independent. \( \square \).
Hypothesis testing

- Suppose \( X_{n \times p} = (X_0, X_1) \) and \( \beta = \left( \begin{array}{c} \beta_0 \\ \beta_1 \end{array} \right) \), where \( \text{rank}(X) = p, \text{rank}(X_0) = p_0 \).

- We want to test \( H_0 : \beta_1 = 0 \) against \( H_1 : \beta_1 \neq 0 \).

- Under \( H_0 \), \( Y = X_0\beta_0 + \varepsilon \).

- Under \( H_0 \), MLEs of \( \beta_0 \) and \( \sigma^2 \) are

\[
\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T Y \\
\hat{\sigma}^2 = \frac{\text{RSS}_0}{n} = \frac{1}{n} (Y - X_0 \hat{\beta}_0)^T (Y - X_0 \hat{\beta}_0)
\]

and these are independent, by Theorem 13.3.

- So fitted values under \( H_0 \) are

\[
\hat{Y} = X_0 (X_0^T X_0)^{-1} X_0^T Y = P_0 Y,
\]

where \( P_0 = X_0 (X_0^T X_0)^{-1} X_0^T \).
Geometric interpretation

\[ U = Xb : b \in \mathbb{R}^p \]

space spanned by cols of \( X \)

\[ \hat{Y} = X\hat{\beta} = PY \]

\[ \sqrt{\text{RSS}} \]

\[ \sqrt{\text{RSS}_0} \]

\[ \sqrt{\text{RSS}_0 - \text{RSS}} \]

\[ \hat{Y} = X_0\hat{\beta}_0 = P_0Y \]

\[ U_0 = X_0b_0 : b_0 \in \mathbb{R}^{p_0} \]

space spanned by cols of \( X_0 \)

\[ H_0 \rightarrow \mathbb{E}(Y) \in U_0 \]
Generalised likelihood ratio test

- The generalised likelihood ratio test of $H_0$ against $H_1$ is

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

$$= \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{RSS_0}{RSS}\right)^{\frac{n}{2}} = \left(1 + \frac{RSS_0 - RSS}{RSS}\right)^{\frac{n}{2}}$$

- We reject $H_0$ when $2 \log \Lambda$ is large, equivalently when $\frac{RSS_0 - RSS}{RSS}$ is large.
- Using the results in Lecture 8, under $H_0$

$$2 \log \Lambda = n \log \left(1 + \frac{RSS_0 - RSS}{RSS}\right)$$

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,
  
  $$\mathbf{Y}^T (I_n - P) \mathbf{Y} = (\mathbf{Y} - \mathbf{X}_0 \beta_0)^T (I_n - P) (\mathbf{Y} - \mathbf{X}_0 \beta_0) \text{ since } (I_n - P) \mathbf{X}_0 = 0,$$
  
  $$\mathbf{Y}^T (P - P_0) \mathbf{Y} = (\mathbf{Y} - \mathbf{X}_0 \beta_0)^T (P - P_0) (\mathbf{Y} - \mathbf{X}_0 \beta_0) \text{ since } (P - P_0) \mathbf{X}_0 = 0,$$
Applying Lemmas 13.2 ($Z^T A_i Z \sim \sigma^2 \chi_i^2$) and 15.1 to $Z = Y - X_0 \beta_0$, $A_1 = I_n - P$, $A_2 = P - P_0$ to get that under $H_0$,

$$
\text{RSS} = Y^T (I_n - P) Y \sim \chi_{n-p}^2
$$

$$
\text{RSS}_0 - \text{RSS} = Y^T (P - P_0) Y \sim \chi_{p-p_0}^2
$$

and these rvs are independent.

So under $H_0$,

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

Hence we reject $H_0$ if $F > F_{p-p_0, n-p}(\alpha)$.

$\text{RSS}_0 - \text{RSS}$ is the 'reduction in the sum of squares due to fitting $\beta_1$.}
Arrangement as an 'analysis of variance' table

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>degrees of freedom (df)</th>
<th>sum of squares</th>
<th>mean square</th>
<th>F statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitted model</td>
<td>$p - p_0$</td>
<td>$RSS_0 - RSS$</td>
<td>$\frac{(RSS_0 - RSS)}{(p - p_0)}$</td>
<td>$\frac{(RSS_0 - RSS)}{(RSS_0 / (n - p))}$</td>
</tr>
<tr>
<td>Residual</td>
<td>$n - p$</td>
<td>$RSS$</td>
<td>$\frac{RSS}{(n - p)}$</td>
<td></td>
</tr>
</tbody>
</table>

The ratio $\frac{(RSS_0 - RSS)}{RSS_0}$ is sometimes known as the *proportion of variance explained by* $\beta_1$, and denoted $R^2$. 
Simple linear regression

- We assume that

\[ Y_i = a' + b(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, \ldots, n, \]

where \( \bar{x} = \frac{\sum x_i}{n} \), and \( \varepsilon_i, i = 1, \ldots, n \) are iid \( \text{N}(0, \sigma^2) \).

- 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 \( Y_i \sim \text{N}(a, \sigma^2) \), 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

\[ \text{RSS}_0 - \text{RSS} = \sum_i \left( (y_i - \overline{y})^2 - (y_i - \overline{y} - \hat{b}(x_i - \bar{x}))^2 \right) = \hat{b}^2 (x_i - \bar{x})^2 = \hat{b}^2 S_{xx}. \]
### 15. Hypothesis testing in the linear model

#### 15.7. Simple linear regression

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>sum of squares</th>
<th>mean square</th>
<th>( F ) statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitted model</td>
<td>1</td>
<td>( \text{RSS}<em>0 - \text{RSS} = \hat{b}^2 S</em>{xx} )</td>
<td>( \hat{b}^2 S_{xx} )</td>
<td>( F = \hat{b}^2 S_{xx} / \tilde{\sigma}^2 )</td>
</tr>
<tr>
<td>Residual</td>
<td>( n - 2 )</td>
<td>( \text{RSS} = \sum_i (y_i - \hat{y})^2 )</td>
<td>( \tilde{\sigma}^2 )</td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{RSS}_0 = \sum_i (y_i - \bar{y})^2
\]

- 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 \( r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \).

- From lecture 14, slide 5, we see that under \( H_0 \), \( \frac{\hat{b}}{\text{s.e.}(\hat{b})} \sim t_{n-2} \), where \( \text{s.e.}(\hat{b}) = \tilde{\sigma} / \sqrt{S_{xx}}. \)

\[
\text{s.e.}(\hat{b}) = \frac{\hat{b} \sqrt{S_{xx}}}{\tilde{\sigma}} = t.
\]

- 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

```r
> 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 \( \mu_i \)'s are unknown constants.

- Fitting this model gives
  \[ \text{RSS} = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{i,j} - \hat{\mu}_i)^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{i,j} - \bar{Y}_i)^2 \]
  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 \( \hat{\mu} = \bar{Y}_.. \), and the fitted values are \( \hat{Y}_{i,j} = \bar{Y}_.. \).

- The observed \( \text{RSS}_0 \) is therefore
  \[ \text{RSS}_0 = \sum_{i} \sum_{j} (y_{i,j} - \bar{y}_{..})^2. \]
The fitted sum of squares is therefore

$$RSS_0 - RSS = \sum_i \sum_j ((y_{i,j} - \bar{y}_{..})^2 - (y_{i,j} - \bar{y}_i)^2) = J \sum_i (\bar{y}_i - \bar{y}_{..})^2.$$  

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>sum of squares</th>
<th>mean square</th>
<th>F statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitted model</td>
<td>$l - 1$</td>
<td>$J \sum_i (\bar{y}<em>i - \bar{y}</em>{..})^2$</td>
<td>$\frac{J \sum_i (\bar{y}<em>i - \bar{y}</em>{..})^2}{(l-1)}$</td>
<td>$F = \frac{J \sum_i (\bar{y}<em>i - \bar{y}</em>{..})^2}{(l-1)\tilde{\sigma}^2}$</td>
</tr>
<tr>
<td>Residual</td>
<td>$n - l$</td>
<td>$\sum_i \sum_j (y_{i,j} - \bar{y}_i)^2$</td>
<td>$\tilde{\sigma}^2$</td>
<td>$\tilde{\sigma}^2$</td>
</tr>
</tbody>
</table>

$$n - 1 \quad \sum_i \sum_j (y_{i,j} - \bar{y}_{..})^2$$
**Example 13.1**

As R code

```r
> summary.aov(fit)

Df  Sum Sq  Mean Sq  F value Pr(>F)
  x       4  507.9  127.00    1.17  0.354
Residuals 20 2170.1  108.50
```

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