Lecture 15. Hypothesis testing in the linear model

15.1. Preliminary lemma

Lemma 15.1

Suppose $Z \sim N_n(0, \sigma^2 I_n)$ 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 $W_1 W_2 = A_1 A_2 Z$, where $A_2 = (A_1 A_2)$. By Proposition 11.1(i), $W_1 \sim N_2 n((0 0), \sigma^2 (A_1 0 0 A_2))$ 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. $\square$.

15.2. Hypothesis testing

Suppose $X_{n \times p} = (X_0 \ X_1)$ and $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, 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 $H_0$, $Y = X_0 \beta_0 + \epsilon$.

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

15.3. Geometric interpretation

![Geometric interpretation diagram]

The space spanned by the columns of $X$ is $U = Xb : b \in \mathbb{R}^p$.
Generalised likelihood ratio test

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

$$\Lambda(Y_0, H_1) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}(Y - X\hat{\beta})^T(Y - X\hat{\beta})\right)$$

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

- 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 $RSS = Y^T(I_n - P)Y$ (see proof of Theorem 13.3 (ii)), and so

$$RSS_0 - RSS = Y^T(I_n - P_0)Y - Y^T(I_n - P)Y = Y^T(P - P_0)Y.$$

- 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

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

- Finally,

$$Y^T(I_n - P)Y = (Y - X_0\hat{\beta})^T(I_n - P)(Y - X_0\beta_0)$$

since $(I_n - P)X_0 = 0,$

and

$$Y^T(P - P_0)Y = (Y - X_0\hat{\beta})^T(P - P_0)(Y - X_0\beta_0)$$

since $(P - P_0)X_0 = 0.$

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>$\frac{(RSS_0 - RSS)/(p - p_0)}{RSS/(n - p)}$</td>
</tr>
<tr>
<td>Residual</td>
<td>$n - p$</td>
<td>$RSS$</td>
<td>$RSS/(n - p)$</td>
<td>$RSS/(n - p)$</td>
</tr>
</tbody>
</table>

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

- We assume that
  \[ Y_i = \alpha + \beta (x_i - \bar{x}) + \epsilon_i, \quad i = 1, \ldots, n, \]
  where \( \bar{x} = \frac{1}{n} \sum x_i/n, \) and \( \epsilon_i, i = 1, \ldots, n \) are iid \( N(0, \sigma^2). \)
- Suppose we want to test the hypothesis \( H_0: \beta = 0, \) 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 \( Y_i \sim N(\alpha, \sigma^2), \) and so \( \bar{Y} = \bar{Y}. \)
- The observed RSS is therefore
  \[ RSS = \sum_i (y_i - \bar{Y})^2 = \sum_i y_i^2 - \sum_i b_i x_i y_i + \sum_i \bar{Y}^2 = \sum_i (y_i - \bar{Y})^2. \]
- The fitted sum of squares is therefore
  \[ RSS_0 = \sum_i (y_i - \bar{Y})^2 = \sum_i y_i^2 - \sum_i b_i x_i y_i + \sum_i \bar{Y}^2 = \sum_i (y_i - \bar{Y})^2. \]

Example 12.1 continued

As R code
\begin{verbatim}
> fit=lm(time~ oxy.s )
> summary.aov(fit)

                 Df Sum Sq Mean Sq F value Pr(>F)
 oxy.s           1 129690 129690  4.198   1.6e-06 ***
 Residuals 22   67968     3089
 ---
 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
\end{verbatim}

Note that the \( F \) statistic, 41.98, is \( -4.68^2, \) the square of the \( t \) statistic on Slide 5 in Lecture 14.

---

Fitted model
\[
\begin{array}{llll}
\text{Source of variation} & \text{d.f.} & \text{sum of squares} & \text{mean square} & F \text{ statistic} \\
\hline
\text{Fitted model} & 1 & RSS_0 - RSS = \hat{b}^2 S_{xx} & \hat{b}^2 S_{xx} & F = \frac{\hat{b}^2 S_{xx}}{\hat{\sigma}^2} \\
\text{Residual} & n - 2 & RSS = \sum_i (y_i - \hat{Y}_i)^2 & \hat{\sigma}^2 & \\
\text{Residual} & n - 1 & RSS_0 = \sum_i (y_i - \bar{Y})^2 & & \\
\end{array}
\]

Note that the proportion of variance explained is \( \hat{b}^2 S_{xx} / S_{yy} = \frac{s_{yy}^2}{s_{xx} s_{yy}} = r^2, \)
where \( r \) is Pearson’s Product Moment Correlation coefficient.

\[ t = \frac{\hat{b}}{s.e.(\hat{b})} = \frac{\hat{b} \sqrt{s.x}}{s.} = t. \]

Checking whether \( |t| > t_{n-2}(\alpha) \) 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.

---

One way analysis of variance with equal numbers in each group

- Assume \( J \) measurements taken in each of \( I \) groups, and that
  \[ Y_{i,j} = \mu_i + \epsilon_{i,j}, \]
  where \( \epsilon_{i,j} \) are independent \( N(0, \sigma^2) \) random variables, and the \( \mu_i \)’s are unknown constants.
- Fitting this model gives
  \[ RSS = \sum_{i=1}^I \sum_{j=1}^I (Y_{i,j} - \hat{\mu})^2 = \sum_{i=1}^I \sum_{j=1}^I (Y_{i,j} - \bar{Y}_i)^2 = n - I \text{ degrees of freedom}. \]
- 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 \( \bar{Y} = \bar{Y}_i, \) and the fitted values are \( \hat{Y}_{i,j} = \bar{Y}_i. \)
- The observed RSS is therefore
  \[ RSS_0 = \sum_i \sum_j (Y_{i,j} - \bar{Y}_i)^2. \]
The fitted sum of squares is therefore

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

<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>$I - 1$</td>
<td>$J \sum_i (\bar{y}<em>{i,j} - \bar{y}</em>{..})^2$</td>
<td>$\frac{J \sum_i (\bar{y}<em>{i,j} - \bar{y}</em>{..})^2}{(I-1)\hat{\sigma}^2}$</td>
<td>$F = \frac{J \sum_i (\bar{y}<em>{i,j} - \bar{y}</em>{..})^2}{(I-1)\hat{\sigma}^2}$</td>
</tr>
<tr>
<td>Residual</td>
<td>$n - I$</td>
<td>$\sum_i \sum_j (y_{i,j} - \bar{y}_{..})^2$</td>
<td>$\hat{\sigma}^2$</td>
<td></td>
</tr>
</tbody>
</table>

Example 13.1
As R code

```r
> 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.

Lecture 15. Hypothesis testing in the linear model 14 (1–14)