

8.7 Likelihood ratio tests

Rep: Model $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$ $\vec{\varepsilon} \sim \text{MVN}(\vec{0}, \sigma^2 I)$

observe $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Likelihood for $\vec{\beta}, \sigma^2$

$$L(\vec{\beta}, \sigma^2 | \vec{y}, X) = f(\vec{y} | X, \vec{\beta}, \sigma^2)$$

density of observed data

Find $\vec{\beta}, \sigma^2$ so that $f(\vec{y})$ fits our data

Maximum likelihood estimators (MLE):

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \vec{y} \quad (\text{least squares})$$

$$\hat{\sigma}^2 = \frac{1}{n} \vec{y}^T (I - H) \vec{y}$$

$$s^2 = \frac{n}{n-k-1} \hat{\sigma}^2$$

(unbiased)

$$H_0: C\vec{\beta} = \vec{0}$$

restricted model

$$H_1: C\vec{\beta} \neq \vec{0}$$

unrestricted model

as above

e.g. $H_0: \vec{\beta}_2 = \vec{0}$

$$H_1: \vec{\beta}_2 \neq \vec{0}$$

$$\vec{\beta} = \begin{pmatrix} \vec{\beta}_1 \\ \vec{\beta}_2 \end{pmatrix}$$

Partial F test

Likelihood ratio:

$$LR = \frac{\max_{H_0} L(\vec{\beta}, \sigma^2)}{\max_{H_1} L(\vec{\beta}, \sigma^2)}$$

$$0 \leq LR \leq 1$$

$\uparrow$
H₁ true

$\uparrow$
H₀ true

NB : Deriving a test from the likelihood ratio principle can in some cases be shown to be optimal.

General result: $-2 \ln(LR) \approx \chi^2(q)$ rank(C)
number
of param
restricted
Under H₀

For us: $C\vec{\beta} = \vec{0}$ F-test gives an exact test derived from the likelihood ratio principle.

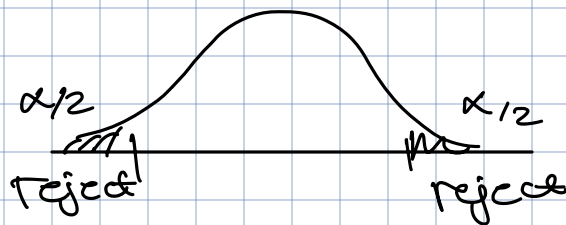
Multiple tests (8.5.2 → Bonferroni)

One test

E.g. $H_0: \beta_j = 0$ $H_1: \beta_j \neq 0$

$$T = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2_{j|j}}}$$

$\overset{H_0}{\sim} t(n-k-1)$



$$P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha \quad (\text{sign. level})$$

E.g. $\alpha = 0.05$

We accept a 5% chance of a false positive result, i.e. rejecting H_0 when H_0 true

k tests

E.g. $H_{0j}: \beta_j = 0$ $H_{1j}: \beta_j \neq 0$

$j = 1, \dots, k$ one test per covariate

Def: 1) comparisonwise (local)

sign. level $\alpha_c = P(\text{reject } H_{0j} \mid H_{0j} \text{ true})$

2) Familywise error rate, α_f :

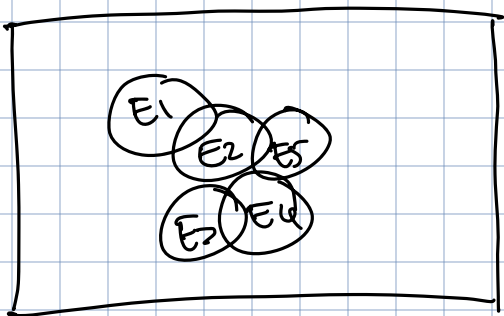
Probability of at least one type I error when all H_0 -s are true.

Aim: control α_f at chosen level α (e.g. $\alpha=0.05$)

Bonferroni method:

Def E_j as event 'reject H_{0j} , H_{0j} true'

$$\alpha_f = P(E_1 \cup E_2 \cup \dots \cup E_k) \leq \sum_{j=1}^k P(E_j)$$



Booles
inequality,

equality if all
events are disjoint.

$P(E_1 \cap E_2) = 0$, strict-

$$\alpha_f \leq \sum_{j=1}^k P(E_j) = k \alpha_c$$

Set $\alpha_c = \frac{\alpha}{k}$, then $\alpha_f \leq \alpha$

Very strict $\rightarrow$ gives conservative tests.

Back to our example
on grocery stores:

$$k=5$$

$$\alpha_f \leq 0.05 \text{ (want)}$$

$$\alpha_c = \frac{0.05}{5} = \underline{\underline{0.01}}$$

7.8 Generalized least squares

Model: $\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$

$E(\vec{\varepsilon}) = \vec{0}$, $\text{cov}(\vec{\varepsilon}) = \sigma^2 V$ (not $\sigma^2 I$)

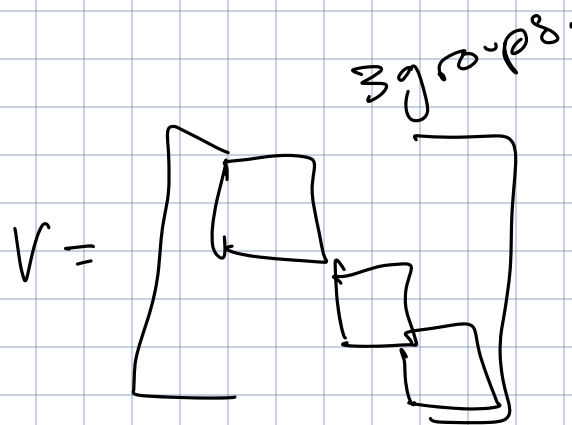
← known for now

Ex 1 Heteroskedasticity



Variance depends on x

Ex 2 Correlated observations



Thm 7.8a

(i) Best linear unbiased estimator

Best = lowest variance

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \vec{Y}$$

← no σ^2

(ii) $\text{cov}(\hat{\beta}) = \sigma^2 (X^T V^{-1} X)^{-1}$

← requires estimate of σ^2 .

(iii) unbiased estimator σ^2 :

$$s^2 = \frac{(\vec{Y} - X\hat{\beta})^T V^{-1} (\vec{Y} - X\hat{\beta})}{n-k-1}$$

NB: we will often not know V , or even that $V \neq I$.

Example 1: Heteroskedasticity

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad x_i \in [-2, 2]$$

$$\varepsilon_i \sim N(0, \sigma^2 \exp(x_i))$$

$$[\beta_0 = 1, \beta_1 = 0.5, \sigma^2 = 1.2^2]$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad V = \text{diag}(e^{x_i})$$

$$\hat{\beta} = \begin{pmatrix} 1.12 \\ 0.52 \end{pmatrix}$$

$$s^2 = 1.46 \quad s = 1.2039 \dots$$

$$\hat{\text{cov}}(\hat{\beta}) = \begin{bmatrix} 0.020 & 0.012 \\ 0.012 & 0.013 \end{bmatrix} \quad \text{Var}(\hat{\beta}_1)$$

R: Built-in methods (GLS)

Model misspecification (Pt. 2) (7.8.2)

$$\hat{\beta}^* = (X^T X)^{-1} X^T \vec{Y} \quad (\text{as before, } \text{cov}(\vec{\epsilon}) = \sigma^2 I)$$

$$E(\hat{\beta}^*) = (X^T X)^{-1} X^T E(\vec{Y}) = (X^T X)^{-1} X^T X \vec{\beta} = \vec{\beta}$$

unbiased 😊

But 'old' expressions for
 $\text{cov}(\hat{\beta})$ and s^2 not correct,
So any other inference (tests, intervals)
not valid.

[R example again]

$$\hat{\beta}^* = \begin{bmatrix} 1.08 \\ 0.40 \end{bmatrix} \quad \leftarrow \text{we can 'trust' this}$$

Example 2

Y : test result

X : # hours studied

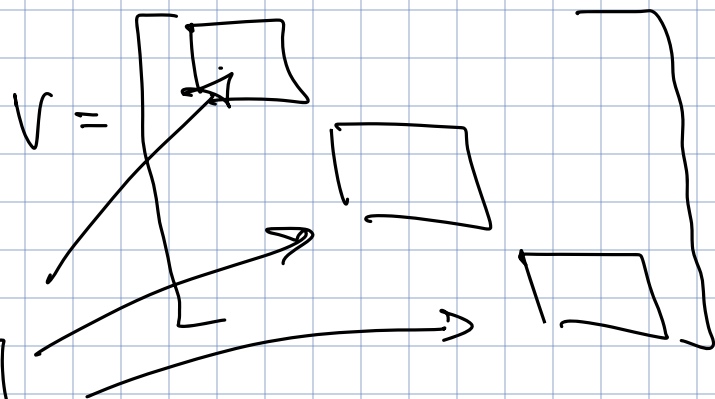
$m = 3$ groups (study programs)

- Design : # hours randomly assigned
- Assume : test results correlated within programs (study groups etc.)

$$\text{Model: } \vec{Y} = X \vec{\beta} + \vec{\epsilon}$$

$$\text{cov}(\vec{\epsilon}) = \sigma^2 V$$

$$\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \dots & \dots & 1 \end{bmatrix}$$



ρ : correlation between test result of two students in same programme.

$$n_1 = 80$$

$$n_2 = 90$$

$$n_3 = 70$$

$$n = n_1 + n_2 + n_3 = 240$$

$$\beta_0 = 40 \quad \beta_1 = 5.5 \quad \sigma^2 = 10^2 \quad \rho = 0.35$$

$$\hat{\beta} = \begin{pmatrix} 35.7 \\ 6.14 \end{pmatrix}$$

$$s = 10.08$$

[manually +
built in]

* unknown ρ :

$$\hat{\rho} = 0.45$$

$$\hat{\beta} = \begin{pmatrix} 35.7 \\ 6.14 \end{pmatrix}$$

$$s^2 = 10.96$$

* OLS

$$\hat{\beta} = \begin{pmatrix} 36.05 \\ 6.18 \end{pmatrix}$$

Linear mixed models (correlated observations)

Example : Y : test result

X : # hours studied

m groups $\begin{cases} \text{study program A} \\ \text{study program B} \\ \text{etc.} \end{cases}$

- Design : # hours randomly assigned
- Assume : test results correlated within programs (study groups etc.)

Model : $Y_{ij} = \beta_0 + \beta_1 X_{ij} + \alpha_i + \epsilon_{ij}$

$i = 1, \dots, m$ group
 $j = 1, \dots, n_i$ subject j in group i

$\beta_0 + \beta_1 X_{ij}$ fixed linear term
 α_i group effect, random
 ϵ_{ij} random variation

$$\alpha_i \sim N(0, \tau_i^2) \quad i = 1, \dots, m$$

$$\epsilon_{ij} \sim N(0, \sigma^2) \quad j = 1, \dots, n_i, \quad i = 1, \dots, m$$

All α -s and ϵ -s independent

$$\text{Var}(Y_{ij}) = \tau_i^2 + \sigma^2$$

$$\text{Cov}(Y_{ij}, Y_{ij'}) = \text{Cov}(\alpha_i + \varepsilon_{ij}, \alpha_i + \varepsilon_{ij'})$$

2 students
in same group

$$= \text{Cov}(\alpha_i, \alpha_i) + \text{Cov}(\varepsilon_{ij}, \alpha_i)$$

$$+ \text{Cov}(\alpha_i, \varepsilon_{ij'}) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'})$$

$$= \text{Var}(\alpha_i, \alpha_i) = \tau_i^2$$

2 students in
different groups

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = 0$$