TMA4267 L13
07.03.2017

PART 3: HYPOTHESIS TESTING AND ANALYSIS OF VARIANCE (ANOVA)

4 lectures + 1 RecEx + 1 CompulsoryEx

Hypothesis testing in linear regression [F.3.3]

$Y = X\beta + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I)$

So far we have looked at two types of hypotheses:

1) Test for significance of one β_j (Ex: β_{money} happiness)

$H_0: \beta_j = 0 \quad$ vs. $\quad H_1: \beta_j \neq 0$

$\Rightarrow$ summary(lm-model) automatically added.

2) Is the regression significant?

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad$ vs $\quad H_1$: at least one $\neq 0$

In addition we migth want to

3) Test of equality (Ex Munich rent: top av good location)

$H_0: \beta_j - \beta_r = 0$ vs $H_1: \beta_j - \beta_r \neq 0$

All of these situations can be written as a general linear hypothesis

$H_0: C\beta = d$ vs $H_1: C\beta \neq d$

C: $r \times p$ matrix

d: $r \times 1$ vector of constants

r linearly independent constraints under H_0

$\text{rank}(C) = r \leq p$

restricted model

model B

unrestricted model

model A

Model B is a subset of model A.

Ex: Happiness, find C:

$\beta_0\ \beta_1\ \beta_2\ \beta_3\ \beta_4$

1) Is there a linear effect of money?

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

$\underset{1\times 5}{C} = [0\ 1\ 0\ 0\ 0]$ $\quad \underset{1\times 1}{d} = 0$

$r = 1$

$C\beta = d \Leftrightarrow \beta_1 = 0$

2) Is the regression significant?

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs H_1: at least one $\neq 0$

$$\underset{4\times 5}{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad d = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$r = 4$

3) Is there a linear effect of money and/or sex?

$H_0: \beta_1 = \beta_2 = 0$ vs H_1: at least one $\neq 0$

$$\underset{2\times 5}{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$r = 2$

Procedure (for testing linear hypotheses)

Urestricted model vs Restricted model

$Y = X\beta + \varepsilon$, $\varepsilon \sim N_n(0, \sigma^2)$ $\nearrow$ $C\beta = d$

Ex: "$\beta_1 = 0$" money example

(A) Unrestricted: fit all covariates: x_1, x_2, x_3, x_4

(B) Restricted: fit only: x_2, x_3, x_4

i) Fit the unrestricted model (A) and compute $SSE = \hat{\varepsilon}^T \hat{\varepsilon}$. Assume p regr. param. fitted.

ii) Fit the restricted model (B) and compute $SSE_{H_0} = \hat{\varepsilon}_{H_0}^T \hat{\varepsilon}_{H_0}$

NB: the restricted model needs to be nested within the unrestricted.

iii) Calculate the test statistic:

$$F_{obs} = \frac{\frac{1}{r}\Delta SSE}{\frac{1}{n-p} SSE} = \frac{\frac{1}{r}\overbrace{(SSE_{H_0} - SSE)}^{\geq 0}}{\frac{1}{n-p} SSE}$$

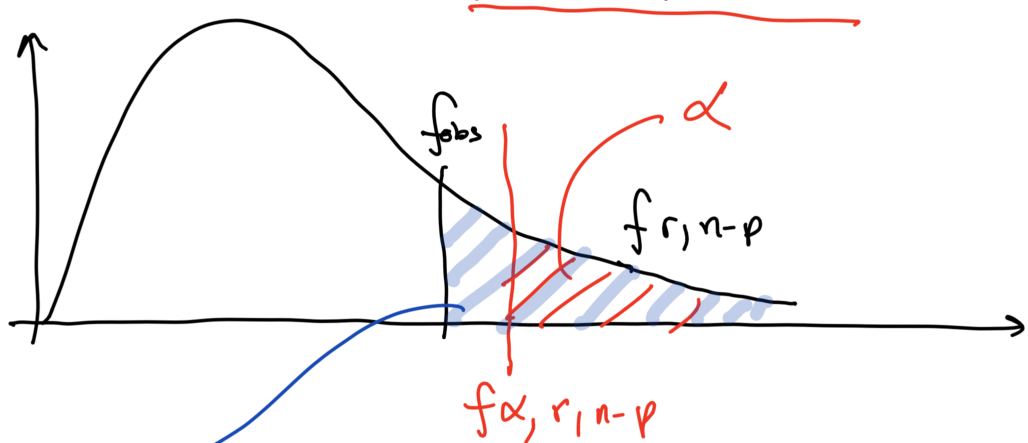
Q: What is the relationship between SSE_{H_0} and SSE?

$$SSE \le SSE_{H_0}$$

$\uparrow$ unrestricted = larger model

iv) Under H_0: $F_{obs} \sim F_{r, n-p}$

Reject H_0 when $f_{obs} > f_{\alpha, r, n-p}$



p-value: $P(F_{r, n-p} > f_{obs})$

How to use this procedure? How to find F_{obs}?

a) Math formula for F_{obs} based on X, C, d, and $\hat{\beta}$, $\hat{\sigma}^2$ from the unrestricted model (A)

on Tuesday!

b) If possible: fit unrestricted model and restricted model and read off SSE to get F_{obs}.

Ex: Happiness: $H_0: \beta_1 = \beta_2 = 0$ $\qquad \hat{\sigma}^2 = \frac{SSE}{n-p}$

A: Full model: $x_1\ x_2\ x_3\ x_4$

$$SSE = \hat{\sigma}^2 \cdot (n-p) = (1.058)^2 \cdot 34 = 38.087$$

B: Restricted model: $x_3\ x_4$

$$SSE_{H_0} = (1.08)^2 \cdot 36 = 41.952$$

$$F_{obs} = \frac{\frac{1}{2}(41.952 - 38.087)}{\frac{1}{34} 38.087} = 1.752$$

$$\text{p-value} = P(F_{2,34} > 1.752) = 0.1934$$

$\Rightarrow$ do not reject H_0: we prefer the smaller model "$x_3 + x_4$".

$H_0: \beta_1 = \beta_2 = 0 \qquad H_1$: at least one $\neq 0$