31 Oct - LEC 7

TEST about a SUBSET OF $\underline{\beta}$

Consider the model (for $i = 1, \dots, n$)

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p_0} x_{ip_0} + \boxed{\beta_{p_0+1} x_{i,p_0+1} + \dots + \beta_p x_{ip}} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

we want to test jointly $(\beta_{p_0+1}, \dots, \beta_p) = \underline{0}$

$$\begin{cases} H_0: \beta_{p_0+1} = \dots = \beta_p = 0 \\ H_1: \overline{H_0} \quad \text{at least one of them is } \neq 0 \quad (\exists r \in \{p_0+1, \dots, p\} : \beta_r \neq 0) \end{cases}$$

Preliminary considerations:

- Under H_1

We have p covariates

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p_0} x_{ip_0} + \beta_{p_0+1} x_{i p_0+1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

We call it the "full model".

When we estimate the model, we obtain:

- estimate $\hat{\underline{\beta}}$ (p-dim. vector)
- residuals $\underline{e} = \underline{y} - X\hat{\underline{\beta}}$
- sum of squared residuals $\underline{e}^T\underline{e}$
- estimate of σ^2, $\hat{\sigma}^2 = \frac{1}{n}\underline{e}^T\underline{e}$. Distribution of the estimator $\frac{n\hat{\Sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$

- Under H_0

We have a model with $p_0 < p$ covariates

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p_0} x_{ip_0} + \varepsilon_i$$

we call it the "restricted model"

We are constraining the coefficients $(\beta_{p_0+1}, \dots, \beta_p)$ to be equal to zero.

When we estimate the model, we obtain:

- estimate $\tilde{\underline{\beta}}$ (p_0-dim. vector)
- residuals $\tilde{\underline{e}} = \underline{y} - X\tilde{\underline{\beta}}$
- sum of squared residuals $\tilde{\underline{e}}^T\tilde{\underline{e}}$
- estimate of σ^2, $\tilde{\sigma}^2 = \frac{1}{n}\tilde{\underline{e}}^T\tilde{\underline{e}}$. Distribution of the estimator $\frac{n\tilde{\Sigma}^2}{\sigma^2} \sim \chi^2_{n-p_0}$

Remark:

The test about a subset of parameters is a test for COMPARING TWO MODELS.
Notice that the two models are NESTED, meaning that the model under H_0 is included into the model under H_1 (it can be obtained from the full model using a set of constraints).
If the models are not nested you can not use this test to compare them.

How we test the hypothesis:

It is useful to write the model in a way to highlight the separation between the unconstrained parameters and the ones we are testing.
First, we formulate the model so that the parameters to test are THE LAST p_0 (simply sort the covariates)
Then, we write

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p_0} \\ \hline \beta_{p_0+1} \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \underline{\beta}^{(0)} \\ \hline \underline{\beta}^{(1)} \end{bmatrix} \qquad \begin{matrix} \underline{\beta}^{(0)} \in \mathbb{R}^{p_0} \\ \underline{\beta}^{(1)} \in \mathbb{R}^{p-p_0} \end{matrix}$$

$\rightarrow$ the system of hypothesis becomes

$$\begin{cases} H_0: \underline{\beta}^{(1)} = \underline{0} \\ H_1: \underline{\beta}^{(1)} \neq \underline{0} \end{cases}$$

Similarly, we write the matrix X as the juxtaposition of two submatrices

$$X = \left[\begin{array}{cccc|ccc} x_{11} & x_{12} & \cdots & x_{1p_0} & x_{1,p_0+1} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np_0} & x_{n,p_0+1} & \cdots & x_{np} \end{array}\right] = \left[\underset{n \times p_0}{X^{(0)}} \,\middle|\, \underset{n \times (p-p_0)}{X^{(1)}}\right]$$

Hence we obtain

FULL MODEL (H_1)

$$\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I)$$

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} = [X^{(0)} X^{(1)}]\begin{bmatrix} \underline{\beta}^{(0)} \\ \underline{\beta}^{(1)} \end{bmatrix} + \underline{\varepsilon} = X^{(0)}\underline{\beta}^{(0)} + X^{(1)}\underline{\beta}^{(1)} + \underline{\varepsilon}$$

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\underline{\beta}}^{(0)} \\ \hat{\underline{\beta}}^{(1)} \end{bmatrix} = (X^TX)^{-1}X^T\underline{y}$$

RESTRICTED MODEL (H_0)

$$\underline{Y} \sim N_n(X^{(0)}\underline{\beta}^{(0)}, \sigma^2 I)$$

$$\underline{Y} = X^{(0)}\underline{\beta}^{(0)} + \underline{\varepsilon}$$

$$\tilde{\underline{\beta}}^{(0)} = (X^{(0)T}X^{(0)})^{-1} X^{(0)T}\underline{y}$$

We know that $\tilde{\underline{e}}^T\tilde{\underline{e}} \geq \underline{e}^T\underline{e}$, since the model under H_0 is a constrained version of the full model.
In particular, the difference between the two will be large if the coefficients that I have forced to zero are actually relevant for the analysis.

If H_0 is true, removing $\underline{\beta}^{(1)}$ in the model will not make a big difference for predicting y.
under H_0, I expect $\tilde{\underline{e}}^T\tilde{\underline{e}} \approx \underline{e}^T\underline{e}$

$$\Rightarrow \frac{\tilde{\underline{e}}^T\tilde{\underline{e}}}{\underline{e}^T\underline{e}} \approx 1 \Rightarrow \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \approx 1 \Rightarrow \frac{SSE_{H_0}}{SSE_{H_1}} \approx 1 \Rightarrow \frac{SSE_{H_0}}{SSE_{H_1}} - 1 \approx 0$$

If H_0 is not true, removing $\underline{\beta}^{(1)}$ will lead to worse results (larger errors).
under H_1, I expect $\tilde{\underline{e}}^T\tilde{\underline{e}} \gg \underline{e}^T\underline{e}$

$$\Rightarrow \frac{\tilde{\underline{e}}^T\tilde{\underline{e}}}{\underline{e}^T\underline{e}} \gg 1 \Rightarrow \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \gg 1 \Rightarrow \frac{SSE_{H_0}}{SSE_{H_1}} \gg 1 \Rightarrow \frac{SSE_{H_0}}{SSE_{H_1}} - 1 \gg 0$$

To perform the test, we are going to use again a function of $\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} - 1$

TEST STATISTIC and DISTRIBUTION

$$F = \frac{\frac{\tilde{\Sigma}^2 - \hat{\Sigma}^2}{p-p_0}}{\frac{\hat{\Sigma}^2}{n-p}} \overset{H_0}{\sim} F_{p-p_0, n-p}$$

analogous formulations

$$F = \frac{\tilde{\Sigma}^2 - \hat{\Sigma}^2}{\hat{\Sigma}^2} \cdot \frac{n-p}{p-p_0} = \frac{\tilde{\underline{E}}^T\tilde{\underline{E}} - \underline{E}^T\underline{E}}{\underline{E}^T\underline{E}} \cdot \frac{n-p}{p-p_0} = \frac{SSE_{H_0} - SSE_{H_1}}{SSE_{H_1}} \cdot \frac{n-p}{p-p_0} \overset{H_0}{\sim} F_{p-p_0, n-p}$$

Note to remember the degrees of freedom

$\tilde{\Sigma}^2 \sim \chi^2_{n-p_0}$

$\hat{\Sigma}^2 \sim \chi^2_{n-p}$

$$\frac{\text{difference of the estimators}}{\text{difference of the d.o.f.}} = \frac{\tilde{\Sigma}^2 - \hat{\Sigma}^2}{(n-p_0)-(n-p)} = \quad n - p_0 - n + p = p - p_0$$

$$F = \frac{\frac{\tilde{\Sigma}^2 - \hat{\Sigma}^2}{p-p_0}}{\frac{\hat{\Sigma}^2}{n-p}} \overset{H_0}{\sim} F_{p-p_0, n-p}$$

$$\frac{\hat{\Sigma}^2}{\text{its d.o.f.}}$$

With the data, we compute the observed value of the test, $f^{obs} = \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \cdot \frac{n-p}{p-p_0}$

How do we define the reject and acceptance regions?

- acceptance region (values that suggest that the data support H_0)
  under H_0: $\tilde{\sigma}^2 \approx \hat{\sigma}^2 \Rightarrow f^{obs} \approx 0$
- reject region (values that suggest that the data are against H_0)
  under H_1: $\tilde{\sigma}^2 \gg \hat{\sigma}^2 \Rightarrow f^{obs} \gg 0$ I reject H_0 for large values of f^{obs}

hence $\mathcal{A} = (0, k)$ and $\mathcal{R} = (k, +\infty)$

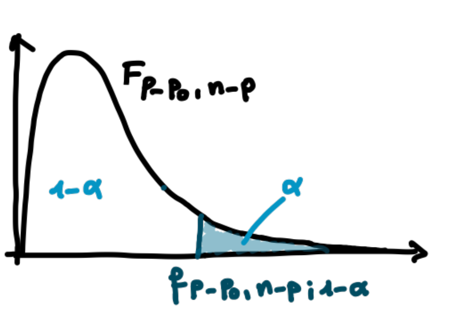
1) FIXED SIGNIFICANCE α

If we fix the significance α, k will be the quantile of level $(1-\alpha)$ of an $F_{p-p_0, n-p}$ distribution

$$\mathcal{R} = (f_{p-p_0, n-p; 1-\alpha} \; ; \; +\infty)$$

with the data: I can compute f^{obs}

- reject H_0 if $f^{obs} > f_{p-p_0, n-p; 1-\alpha}$



2) P-VALUE

Alternatively, the p-value is $\alpha^{obs} = \mathbb{P}_{H_0}(F > f^{obs})$