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

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

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

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

idea:

under H_1, I have p covariates (we call it the "full model")

I can estimate this model, obtaining $\hat{\underline{\beta}}$, and compute the residuals $\underline{e} = \underline{y} - X\hat{\underline{\beta}}$

then, I compute the sum of the squared residuals $\underline{e}^T\underline{e}$

under H_0, I have the model $Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p_0} x_{ip_0} + \varepsilon_i \quad \to p_0$ covariates

I am constraining the coefficients $(\beta_{p_0+1}, \dots, \beta_p)$ to be $= 0$ (we call it the "restricted model")

I can estimate this model, obtaining $\tilde{\underline{\beta}}$, and I obtain the residuals $\tilde{\underline{e}} = \underline{y} - \tilde{\underline{y}}$

The sum of squared residuals is $\tilde{\underline{e}}^T\tilde{\underline{e}}$

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.

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).
Careful: if the models are not nested you can not use the 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.
Write

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

$\to$ 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 partition the matrix X into 2 sub-matrices

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

Hence we obtain

FULL MODEL (H_1)	RESTRICTED MODEL (H_0)
$\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I)$	$\underline{Y} \sim N_n(X^{(0)}\underline{\beta}^{(0)}, \sigma^2 I)$
$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} = [X^{(0)} X^{(1)}] \begin{bmatrix} \underline{\beta}^{(0)} \\ \underline{\beta}^{(1)} \end{bmatrix} + \underline{\varepsilon} = X^{(0)}\underline{\beta}^{(0)} + X^{(1)}\underline{\beta}^{(1)} + \underline{\varepsilon}$	$\underline{Y} = X^{(0)}\underline{\beta}^{(0)} + \underline{\varepsilon}$
$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\underline{\beta}}^{(0)} \\ \hat{\underline{\beta}}^{(1)} \end{bmatrix} = (X^TX)^{-1}X^T\underline{y}$	$\tilde{\underline{\beta}}^{(0)} = (X^{(0)T}X^{(0)})^{-1} X^{(0)T}\underline{y}$

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

$$\Rightarrow \frac{\tilde{\underline{e}}^T\tilde{\underline{e}}}{\underline{e}^T\underline{e}} \approx 1 \Rightarrow \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \approx 1$$

under H_1, I expect $\tilde{\underline{e}}^T\tilde{\underline{e}} \gg \underline{e}^T\underline{e}$

$$\Rightarrow \frac{\tilde{\underline{e}}^T\tilde{\underline{e}}}{\underline{e}^T\underline{e}} \gg 1 \Rightarrow \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \gg 1$$

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

In particular, consider

$$F = \frac{\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{p-p_0}}{\frac{\hat{\sigma}^2}{n-p}} = \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}} \cdot \frac{n-p}{p-p_0} = \frac{\tilde{\underline{E}}^T\tilde{\underline{E}} - \underline{E}^T\underline{E}}{\underline{E}^T\underline{E}} \cdot \frac{n-p}{p-p_0}$$

It holds $F \overset{H_0}{\sim} F_{p-p_0, n-p}$

REMARKS

(1) the degrees of freedom

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

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

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

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

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

(2) the test about an individual coefficient β_j (test 1) and about the overall significance (Test 2) are particular cases of this test.

- **TEST about a SINGLE PARAMETER β_j**

  Assume we are testing the significance of the last parameter β_p.
  (or simply sort the columns of X so that the last covariate is the one corresponding to the parameter of interest)
  Testing β_p is equivalent to testing a model with $p_0 = p-1$ covariates
  In this case we can partition $\underline{\beta}$ and X as

  $$\underline{\beta} = \left[\begin{array}{c} \beta_1 \\ \vdots \\ \beta_{p-1} \\ \hline \beta_p \end{array}\right] \begin{array}{l} p_0 = p-1 \\ \\ \\ 1 \end{array} \qquad X = [\underline{x}_1 \dots \underline{x}_{p-1} \mid \underline{x}_p]$$

  Indeed $p_0 = p-1$ and the test becomes

  $$F = \frac{\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{1}}{\frac{\hat{\sigma}^2}{n-p}} \overset{H_0}{\sim} F_{1, n-p} \qquad F = (T_p)^2 \quad \text{with} \quad T_p = \frac{\hat{\beta}_p - 0}{\sqrt{\hat{V}(\hat{\beta}_p)}} \overset{H_0}{\sim} t_{n-p}$$

  $\downarrow$

  (recall: if $V \sim t_m$, then $V^2 \sim F_{1,m}$)

- **TEST ABOUT THE OVERALL SIGNIFICANCE**

  if we consider $p_0 = 1$

  $$\text{then } \underline{\beta} = \left[\begin{array}{c} \beta_1 \\ \hline \beta_2 \\ \vdots \\ \beta_p \end{array}\right] \begin{array}{l} 1 = p_0 \\ \\ p-1 \\ \\ \end{array}$$

  And H_0 becomes $H_0 : \beta_2 = \beta_3 = \dots = \beta_p = 0$

  $$\text{the test is } F = \frac{\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{p-1}}{\frac{\hat{\sigma}^2}{n-p}} \overset{H_0}{\sim} F_{p-1, n-p}$$

  which is equivalent to the test on the coefficient R^2 ($H_0 : R^2 = 0$ vs $H_1 : R^2 \neq 0$)

(3) Geometric interpretation of the test

Consider again the representation of the model in an n-dimensional space.
Here, the variables $(\underline{y}, \underline{x}_1, \dots, \underline{x}_p)$ are n-dimensional vectors, with coordinates the observations on the n units.
The covariates $(\underline{x}_1, \dots, \underline{x}_p)$ identify a subspace of dimension p, $C(X)$.
This subspace is defined by all linear combinations $\beta_1\underline{x}_1 + \dots + \beta_p\underline{x}_p = X\underline{\beta}$.
The mean of $\underline{Y}$ is $\underline{\mu} = X\underline{\beta} \Rightarrow$ the mean of $\underline{Y}$ belongs to $C(X)$.
The vector $\underline{y}$ in general will not belong to $C(X)$: indeed we have seen that $\hat{\underline{\mu}} = \hat{\underline{y}}$ is the orthogonal projection of $\underline{y}$ onto $C(X)$.

What happens when we compare NESTED models?
example with 2 variables $(\underline{x}_1, \underline{x}_2)$
$C(X)$ is the subspace of $\beta_1\underline{x}_1 + \beta_2\underline{x}_2 \to \hat{\underline{y}}$ is the vector of this space that minimizes the distance between $\underline{y}$ and $X\underline{\beta}$: $\hat{\underline{y}} = \hat{\beta}_1\underline{x}_1 + \hat{\beta}_2\underline{x}_2$
Assume we want to test
$H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$
Under H_0, I am saying that $\tilde{\underline{y}} = \tilde{\beta}_1\underline{x}_1 \to \tilde{\underline{y}}$ will belong to the subspace defined by a straight line (and not to the entire plane)
$\to$ This is a constrained estimate:
$\tilde{\beta}_1$ is the value minimizing the distance between $\underline{y}$ and $X\beta = \beta_1\underline{x}_1$
Test: we look how far is $\hat{\underline{y}}$ to $\tilde{\underline{y}}$ or, equivalently, $\tilde{\underline{e}}$ to $\underline{e}$
i.e. $\underline{d} = \tilde{\underline{e}} - \underline{e}$

example with 2 covariates x_1 and x_2
and I test $\beta_2 = 0$