University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Comparing regression equations

Using two different data sets on the same variables we build the following two regression models

$$\mathbf{y_1} = \mathbf{X_1}\boldsymbol{\beta_1} + \boldsymbol{\epsilon_1}$$
$$\mathbf{y_2} = \mathbf{X_2}\boldsymbol{\beta_2} + \boldsymbol{\epsilon_2}$$

where,
$\mathbf{y_1}$ is $n_1 \times 1$, $\mathbf{y_2}$ is $n_2 \times 1$,
$\mathbf{X_1}$ is $n_1 \times (k+1)$, $\mathbf{X_2}$ is $n_2 \times (k+1)$
$\boldsymbol{\beta_1}$ is $(k+1) \times 1$, $\boldsymbol{\beta_2}$ is $(k+1) \times 1$.

Suppose the first p elements of the vectors $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are the same. Then, we can write the two vectors as

$$\boldsymbol{\beta_1} = \begin{pmatrix} \boldsymbol{\beta^1} \\ \boldsymbol{\beta^2_1} \end{pmatrix} \text{ and } \boldsymbol{\beta_2} = \begin{pmatrix} \boldsymbol{\beta^1} \\ \boldsymbol{\beta^2_2} \end{pmatrix}$$

where,
$\boldsymbol{\beta^1}$ is $p \times 1$, $\boldsymbol{\beta^2_1}$ is $(k+1-p) \times 1$, and $\boldsymbol{\beta^2_2}$ is $(k+1-p) \times 1$.

We wish to test the hypothesis
$H_0 : \boldsymbol{\beta^2_1} - \boldsymbol{\beta^2_2} = \mathbf{0}$
$H_a : \boldsymbol{\beta^2_1} - \boldsymbol{\beta^2_2} \neq \mathbf{0}$.

The test will be performed through the general F test discussed in previous lectures. But first, let's partition the matrices $\mathbf{X_1}$ and $\mathbf{X_2}$ as follows:

$$\mathbf{X_1} = \begin{pmatrix} \mathbf{X^1_1} & \mathbf{X^2_1} \end{pmatrix} \text{ and } \mathbf{X_2} = \begin{pmatrix} \mathbf{X^1_2} & \mathbf{X^2_2} \end{pmatrix}$$

Note: The columns of $\mathbf{X^1_1}$ correspond to $\boldsymbol{\beta^1}$ and the columns of $\mathbf{X^2_1}$ correspond to $\boldsymbol{\beta^2_1}$. Similarly, the columns of $\mathbf{X^1_2}$ correspond also to $\boldsymbol{\beta^1}$ and the columns of $\mathbf{X^2_2}$ correspond to $\boldsymbol{\beta^2_2}$.

The two models can be combined into one regression model:

$$\begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{pmatrix} = \begin{pmatrix} \mathbf{X^1_1} & \mathbf{X^2_1} & \mathbf{0} \\ \mathbf{X^1_2} & \mathbf{0} & \mathbf{X^2_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta^1} \\ \boldsymbol{\beta^2_1} \\ \boldsymbol{\beta^2_2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \end{pmatrix}.$$

This model that combines all the information is of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

and the hypothesis $\boldsymbol{\beta^2_1} - \boldsymbol{\beta^2_2} = \mathbf{0}$ can be tested using
$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$
$H_a : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$
with the general F test

$$F_{m,n-k-1} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})' \left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})}{m s_e^2},$$

where the matrix $\mathbf{C}$ is defines as

$$\mathbf{C} = \begin{pmatrix} \mathbf{0}_{k+1-p,p} & \mathbf{I}_{k+1-p} & -\mathbf{I}_{k+1-p} \end{pmatrix}$$

Therefore the dimensions of $\mathbf{C}$ are $(k+1-p) \times [2(k+1)-p]$.

Example:
Suppose $k = 5$ and we we want to test
$H_0 : \beta_4^1 = \beta_4^2, \beta_5^1 = \beta_5^2$
H_a : Not true

Note:
Subscript ij refers to observation i for variable j.
Superscript 1 or 2 refer to dataset 1 or 2.

This is the formulation:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ y_{22} \\ y_{32} \\ \vdots \\ y_{n_2 2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11}^1 & x_{12}^1 & x_{13}^1 & x_{14}^1 & x_{15}^1 & 0 & 0 \\ 1 & x_{21}^1 & x_{22}^1 & x_{23}^1 & x_{24}^1 & x_{25}^1 & 0 & 0 \\ 1 & x_{31}^1 & x_{32}^1 & x_{33}^1 & x_{34}^1 & x_{35}^1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n_1 1}^1 & x_{n_1 2}^1 & x_{n_1 3}^1 & x_{n_1 4}^1 & x_{n_1 5}^1 & 0 & 0 \\ 1 & x_{11}^2 & x_{12}^2 & x_{13}^2 & 0 & 0 & x_{14}^2 & x_{15}^2 \\ 1 & x_{21}^2 & x_{22}^2 & x_{23}^2 & 0 & 0 & x_{24}^2 & x_{25}^2 \\ 1 & x_{31}^2 & x_{32}^2 & x_{33}^2 & 0 & 0 & x_{34}^2 & x_{35}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n_2 1}^2 & x_{n_2 2}^2 & x_{n_2 3}^2 & 0 & 0 & x_{n_2 4}^2 & x_{n_2 5}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4^1 \\ \beta_5^1 \\ \beta_4^2 \\ \beta_5^2 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{31} \\ \vdots \\ \epsilon_{n_1 1} \\ \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{n_2 1} \end{pmatrix}$$

In this example we have $k = 4, p = 4$ and $\mathbf{C}$ is given by $\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$.