University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

F test for the general linear hypothesis

Consider the regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \epsilon_i, \quad i = 1, \ldots, n.$$

Also, $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$, and $\text{var}(\epsilon_i) = \sigma^2$.

Suppose we want to test the following linear hypotheses:

a. $H_0 : \beta_2 = 0$
   $H_a : \beta_2 \neq 0$

b. $H_0 : \beta_2 = 3$
   $H_a : \beta_2 \neq 3$

c. $H_0 : \beta_1 = \beta_5$ or $\beta_1 - \beta_5 = 0$
   $H_a : \beta_1 \neq \beta_5$ or $\beta_1 - \beta_5 \neq 0$

d. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
   $H_a :$ At least one $\beta_i \neq 0$

   This hypothesis can be expressed as:
   $H_0 : \boldsymbol{\beta}_{(\mathbf{0})} = \mathbf{0}$
   $H_a : \boldsymbol{\beta}_{(\mathbf{0})} \neq \mathbf{0}$
   where, $\boldsymbol{\beta}_{(\mathbf{0})} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5))'$

e. $H_0 : (\beta_2, \beta_5)' = \mathbf{0}$
   $H_a : (\beta_2, \beta_5)' \neq \mathbf{0}$

All these hypotheses above can be expressed through the general linear hypothesis:
$H_0 : \mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma} = \mathbf{0}$
$H_a : \mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma} \neq \mathbf{0}$

Let's find the matrix $\mathbf{C}$ and the vector $\boldsymbol{\gamma}$ for each one of the hypotheses (a)-(e) above:

a. $\mathbf{C} = (0, 0, 1, 0, 0, 0)$ and $\boldsymbol{\gamma} = 0$.

b. $\mathbf{C} = (0, 0, 1, 0, 0, 0)$ and $\boldsymbol{\gamma} = 3$

c. $\mathbf{C} = (0, 1, 0, 0, 0, -1)$ and $\boldsymbol{\gamma} = 0$

d. This is also called the overall significance of the model. The matrix $\mathbf{C}$ and the vector $\boldsymbol{\gamma}$ will be defined as follows:

$$\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma} = \mathbf{0}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

   This will give us the hypothesis $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

e. We are testing here whether the two parameters (β_2, β_5) are significant simultaneously. The matrix $\mathbf{C}$ and the vector $\boldsymbol{\gamma}$ will be defined as follows:

$$\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma} = \mathbf{0}$$

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   This will give us the hypothesis $\beta_2 = \beta_5 = 0$.

Example:
Access the data:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics100C/body_fat.txt", header=TRUE)

# Dependent variable:
a$y

#Independent variables:
x1 <- a$x6
x2 <- a$x7
x3 <- a$x8
x4 <- a$x9
x5 <- a$x10
```

a. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
   H_a : At least one $\beta_i \neq 0$

```
ones <- rep(1, nrow(a))

#Construct the design matrix X:
X <- as.matrix(cbind(ones,x1,x2,x3,x4,x5))

#Estimate the beta vector:
beta_hat <- solve(t(X) %*% X) %*% t(X) %*% a$y

#Define the matrix C:
q <- c(0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1)
C <- matrix(q,5,6,byrow=TRUE)

#Vector gamma:
g <- c(0,0,0,0,0)

#Compute se^2:
se2 <- (t(a$y) %*% a$y - t(beta_hat) %*% t(X) %*% a$y) / (nrow(a)-5-1)

#Compute the F statistics:
F <- (t(C%*%beta_hat-g)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%(C %*%beta_hat-g)) / (5*se2)
```

b. $H_0 : (\beta_2, \beta_5)' = \mathbf{0}$
   $H_a : (\beta_2, \beta_5)' \neq \mathbf{0}$

```
#Define matrix C and vector gamma:
q <- c(0,0,1,0,0,0,0,0,0,0,0,1)
C <- matrix(q, 2,6, byrow=TRUE)

g <- c(0,0)

F <- (t(C%*%beta_hat-g)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%(C%*%beta_hat-g)) / (2*se2)
```

c. $H_0 : \beta_2 = 0$
   $H_a : \beta_2 \neq 0$

```
#Define matrix C and vector gamma:

q <- c(0,0,1,0,0,0)
C <- matrix(q, 1,6, byrow=TRUE)

g <- c(0)

F <- (t(C%*%beta_hat-g)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%(C%*% beta_hat-g)) / (1*se2)
```