7 Testing Multiple Linear Restrictions: The F-test

Suppose the model is specified by

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\
H_0 &: \beta_2 = 0 \text{ and } \beta_3 = 0 \\
H_1 &: H_0 \text{ is not true}
\end{aligned}
$$

→ want test if x_2 and x_3 no impact on y

We can use the F-test to test this type of "multiple hypotheses".

1. Our full model is called the "unrestricted" model (ur). Suppose it can be expressed as: (Big model)

   > $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$ is true

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + u$$

   ⇒ Reject H_0

2. The model which takes out x (which we think its associated $\beta = 0$) is called the restricted model (r). ← small model

   > $y = \beta_0 + \beta_1 x_1 + u$ is true ⇒ do not reject H_0

- Suppose there are "q" number of β that we want to perform joint test of $= 0$

  e.g. $q = 2$.

$H_0: \beta_{k-q+1} = \beta_{k-q+2} = \dots \beta_k = 0$

$H_1: H_0$ is not true

$$\underbrace{\underbrace{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-q} x_{k-q}}_{(r)} + \beta_{k-q+1} x_{k-q+1} + \beta_{k-q+2} x_{k-q+2} \dots + \beta_k x_k + u}_{ur}$$

$$F \equiv \frac{\dfrac{(SSR_r - SSR_{ur})}{q}}{\dfrac{SSR_{ur}}{(n-k-1)}}$$

← Always positive b/c $SSR_{ur} < SSR_r$ b/c more x = better explain

$(n-k-1)$ ← d.f. of "ur" model

*. so, if every time add 1 more x variable, SSR↓, and R^2↑, but if we keep add, every time we add 1 more x $\text{var}(\hat{\beta})$ will increase, making prediction of β less precise.

3. Some useful facts

① $R^2_{ur} > R^2_r$

② include more x, model better explained. However, we would like to reject Ho if the inclusion of extra variable does not improve the model enough

4. Other ways to calculate the F-statistics:

=> from $R^2 = 1 - \frac{SSR \to RSS}{SST \to TSS}$

we have $F = \frac{(R^2_{ur} - R^2_r) / q}{(1 - R^2_{ur}) / (n - k - 1)}$

q ← # of B that are set to 0

n ← number of obs; k ← # of slope B; 1 ← intercept

=> if want to test over all significance of model

$H_0: B_1 = B_2 = B_3 .. B_k = 0$, H_a: other wise

$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$ ← R^2 of the model = ur; the "r" model has no x at all

Example: Suppose we are interested in understanding the determinant of a baseball player's salary.

salary	=	season salary
years	=	years in major leagues
gamesyr	=	games per year in the league
bavg	=	career batting average
hrunsyr	=	homeruns per year
rbisyr	=	runs batted in per year

r: salary, years, gamesyr; ur: all

If want to test whether performance has impact on salary

$H_0: B_{bavg} = B_{hrunsyr} = B_{rbisyr} = 0$

H_a: other wise

- the unrestricted model (ur) is defined by

```
. regress log_salary years gamesyr bavg hrunsyr rbisyr

      Source |       SS           df       MS      Number of obs   =       353
-------------+----------------------------------   F(  5,    347)  =    117.06
       Model |  308.989208         5  61.7978416   Prob > F        =    0.0000
    Residual |  183.186327       347  .527914487   R-squared       =    0.6278
-------------+----------------------------------   Adj R-squared   =    0.6224
       Total |  492.175535       352  1.39822595   Root MSE        =    .72658

------------------------------------------------------------------------------
  log_salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       years |   .0688626   .0121145     5.68   0.000     .0450355    .0926898
     gamesyr |   .0125521   .0026468     4.74   0.000     .0073464    .0177578
        bavg |   .0009786   .0011035     0.89   0.376    -.0011918     .003149
     hrunsyr |   .0144295    .016057     0.90   0.369    -.0171518    .0460107
      rbisyr |   .0107657    .007175     1.50   0.134    -.0033462    .0248776
       _cons |   11.19242   .2888229    38.75   0.000     10.62435    11.76048
------------------------------------------------------------------------------
```

1 2 3 4 5

intercept

so k=5

- the restricted model (r) is defined by

```
. regress log_salary years gamesyr

      Source |       SS           df       MS      Number of obs   =       353
-------------+----------------------------------   F(  2,    350)  =    259.32
       Model |  293.864058         2  146.932029   Prob > F        =    0.0000
    Residual |  198.311477       350  .566604221   R-squared       =    0.5971
-------------+----------------------------------   Adj R-squared   =    0.5948
       Total |  492.175535       352  1.39822595   Root MSE        =    .75273

------------------------------------------------------------------------------
  log_salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       years |    .071318    .012505     5.70   0.000     .0467236    .0959124
     gamesyr |   .0201745   .0013429    15.02   0.000     .0175334    .0228156
       _cons |    11.2238    .108312   103.62   0.000     11.01078    11.43683
------------------------------------------------------------------------------
```

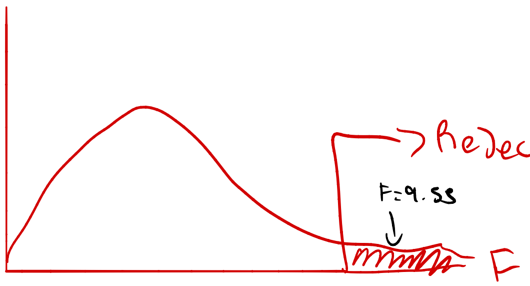
SSE
SSR
SST

Now, our H_0 and H_a becomes

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(198.311 - 183.186)/3}{(183.186/(353-5-1))} \approx 9.55$$

HW:

$$F = \frac{(R^2/q)}{(1-R^2)/n-k-1} = ??$$



→ Reject at 5% ∵ suppose 5% level of sig.

since F = 9.55 > 2.6 we reject H_0 at 5% level and conclude that performance have joint effects on salary.

use ∞ and 3 (page 816) in table $F_{3,\infty}$, $C = 2.6$

8 How the Hypothesis Testing is done in Practice

1. Check the values of $t - statistic$ reported by the statistical software (i.e. STATA, SPSS, SAS)

   $\Rightarrow$ These $t - statistics$ are to test $H_0 : \beta_i = 0$

   z table

   $\Rightarrow$ If the d.f. > 30, then when $t \ > 1.96$, we can reject H_0 with 5% sig. level

   $\Rightarrow$ **When** $t \ \ > \mathbf{1.96}$, we can say that β_i is **statistically significant** at 5% level. (value of $\beta_i \neq 0$)

   $\Rightarrow$ **When** $t \ \ < \mathbf{1.96}$ we can say that β_i is **not statistically significant** at 5% level.

   $\Rightarrow$ If $t \ \ < 1.96$ we can drop x_i from the model

   $\Rightarrow$ After we drop x_i, we estimate the new regression function and obtain a new set of $\hat{\beta}$.

2. We can also perform other hypothesis testings of interest.

   e.g. $H_0 : \beta_i = \beta_j$

   or $\quad H_0 : \beta_i = 5 \quad$ etc.

   or perform an F-test for testing multiple linear restrictions

3. Usually, in economics, the estimation results are reported using this form

sales →

other company performance

ceo characteristic

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
log(*sales*)	.224 (.027)	.158 (.040)	.188 (.040)
log(*mktval*)	——	.112 (.050)	.100 (.049)
profmarg	——	−.0023 (.0022)	−.0022 (.0021)
ceoten	——	——	.0171 (.0055)
comten	——	——	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 0.25)	4.57 (0.25)
Observations *R*-squared	177 .281	177 .304	177 .353