F - test motivation

⇒ we want to test the significance of a group of hypotheses (multiple hypotheses)

$$\text{Grade}_{325} = \beta_0 + \beta_1 \,\#\text{times\_front} + \beta_2 \,\#\text{times\_back} + \beta_3 \text{hr\_study} + \beta_4 \text{past\_GPA} + \beta_5 \text{gender} + u$$

H_0 : seat position doesn't have impact on GPA

$$\beta_1 = 0 \text{ and } \beta_2 = 0 \Rightarrow \beta_1 = \beta_2 = 0$$

H_a : seat position matters

$$\left.\begin{array}{lll} & \beta_1 \neq 0 & \text{and } \beta_2 \neq 0 \\ \text{or} & \beta_1 \neq 0 & \text{and } \beta_2 = 0 \\ \text{or} & \beta_1 = 0 & \text{and } \beta_2 \neq 0 \end{array}\right\} \begin{array}{l} \text{at least one of} \\ \text{the } \beta_1, \beta_2 \neq 0 \end{array}$$

7 Testing Multiple Linear Restrictions: The F-test

Suppose the model is specified by

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

⇒ want to test if X_1 & X_2 both have no impact on y.

H_a,

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

Big model

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

   $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$ is true ⇒ Reject H_0

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

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 would like to perform a joint-test of $= 0$

e.g. in this model $q=2$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{K-q} X_{K-q} + u$$

$$H_0: \beta_{K-q+1} = \beta_{K-q+2} = \ldots = \beta_K = 0$$

(the last q βs $= 0$)

H_0 is not true.

$$Y = \underbrace{\underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{K-q} X_{K-q}}_{(r)} + \beta_{K-q-1} X_{K-q-1} + \beta_{K-q+2} X_{K-q+2} + \ldots + \beta_K X_K + u}_{ur}$$

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

d.f. of the "ur" model

This is always (+) b/c $SSR_{ur} < SSR_r$ Every time you add 1 more X_q the model will be better explained

3. Some useful facts

① $R^2_{ur} > R^2_r$ because any additional X would increase R^2 (improve fit).

$\Rightarrow SSR_{ur} < SSR_r$

② By including more X, the model is certainly better explained. However, we would like to reject H_0 if the conclusion of extra variables does not improve the model enough.

4. Other ways to calculate the F-statistics:

$$\Rightarrow \text{From } R^2 = 1 - \frac{SSR^{\nearrow RSS}}{SST_{\searrow TSS}}$$

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

# of β that are set to "0"

intercept

# of slope β.

# of obs.

⇒ If we want to test the overall significance of the model

$H_0: \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$

H_a : otherwise

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

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.

y *salary* = season salary

$r\ \{\ ur\ \{$ *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

If we want to test whether performance has any impact on salary

$H_0: \beta_{bavg} = \beta_{hrunsyr} = \beta_{rbisyr} = 0$

H_a: otherwise is true

! the unrestricted model (ur) is defined by ur model

y x

```
. 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

HW:

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

$\equiv$?

! the restricted model (r) is defined by

- when considering each of the performance X one-by-one none of them has a significant impact at 5%.

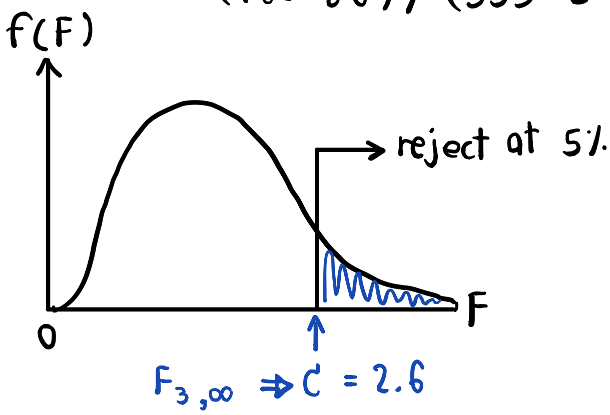
```
. regress log_salary years gamesyr

      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
------------------------------------------------------------------------------
```

Now, our H_0 and H_a becomes

- But when performing an F-test, performances have joint impact.

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

$$\equiv \frac{(198.311 - 183.186)/3}{(183.86)/(353-5-1)} \approx 9.55$$



Let's use 5% level of sig.

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

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)

   # These t " *statistics* are to test $H_0 : \beta_i = 0$

   z-table with 5% sig. level

   # If the d.f. > 30, then when $t \ > 1.96$, we can reject H_0

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

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

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

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

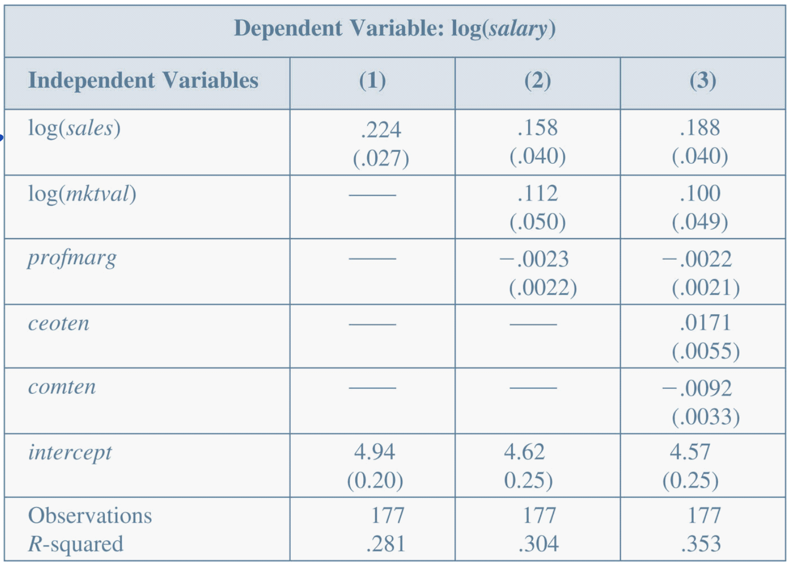
2. We can also perform other hypothesis testings of interest.

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

   or $H_0 : \beta_i = 5$ etc.

   or perform an F-test for testing multiple linear restrictions

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



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

sales

other company performance

CEO characteristics

like a simple regression with 1 X.