Important: see footnote in page 2

Econ 103: Introduction to Econometrics
Week 9

TA: Patrick Molligo *

Winter 2022

Testing Joint Hypotheses

Given an i.i.d. sample $\{(Y_i, X_{2i}, \cdots, X_{Ki})\}_{i=1}^n$ and usual assumptions, consider a model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + e_i$$

With more complex models can we test more complex hypotheses?

- **Joint hypothesis**: a null hypothesis made up of several equalities,
- H_0: equalities are all true together
- H_1: *some* of these equalities are not true → at least one it is not true
- Common example, suppose all (non-constant) regressors have coefficients of zero:

$$H_0 \ : \ \beta_2 = 0 \text{ and } \beta_3 = 0 \cdots \text{ and } \beta_K = 0 \qquad H_1 \ : \ \beta_2 \neq 0 \text{ or } \beta_3 \neq 0 \text{ or } \cdots \text{ or } \beta_K \neq 0$$

- Another example:
  - Collect data on grades, time spent studying, time in lecture, and time in TA sessions
  - Estimate the following model:

$$grade_i = \beta_1 + \beta_2 study_i + \beta_3 lecture_i + \beta_4 TA_i + e_i$$

  - Student claims time in lecture is twice as helpful as time in TA sessions, and that studying on her own is useless. State claim as a null hypothesis against the natural alternative:

$$H_0 \ : \ \beta_2 = 0 \text{ and } \beta_3 = 2\beta_4, \qquad H_1 \ : \ \beta_2 \neq 0 \text{ or } \beta_3 \neq 2\beta_4$$

    Equivalently:

$$H_0 \ : \ \beta_2 = 0 \text{ and } \beta_3 - 2\beta_4 = 0, \qquad H_1 \ : \ \beta_2 \neq 0 \text{ or } \beta_3 - 2\beta_4 \neq 0$$

*I'd like to thank Ben Pirie and Lucas Zhang for providing much of this material. All mistakes are my own.

The F distribution and the p-value

- Recall that the OLS estimates are defined with a minimization problem

$$b_1^U, \cdots, b_K^U \qquad \text{solves} \qquad \min_{b_1^*,\cdots,b_K^*} \sum_{i=1}^{n} (Y_i - b_1^* - b_2^* X_{2i} - \cdots - b_K^* X_{Ki})^2$$

- Refer to the usual OLS estimates as **unrestricted estimates**, denoted b_k^U

- **If** H_0 **true**, then imposing H_0 shouldn't change the value of the minimum much
- Define **restricted estimates** as the solution to a constrained minimization problem:

$$b_1^R, \cdots, b_K^R \qquad \text{solves} \qquad \min_{\substack{b_1^*,\cdots,b_K^* \\ \text{subject to } H_0}} \sum_{i=1}^{n} (Y_i - b_1^* - b_2^* X_{2i} - \cdots - b_K^* X_{Ki})^2$$

constraints

- Define the restricted and unrestricted sum of squared errors:

$$SSE_U = \sum_{i=1}^{n} \left(Y_i - b_1^U - b_2^U X_{2i} - \cdots - b_K^U X_{Ki}\right)^2$$
$$SSE_R = \sum_{i=1}^{n} \left(Y_i - b_1^R - b_2^R X_{2i} - \cdots - b_K^R X_{Ki}\right)^2$$

- Unrestricted estimates minimize $SSE_U \implies$ can't generate a smaller SSE
- Unrestricted minimum is weakly smaller than restricted minimum. Mathematically,

$$SSE_R \geq SSE_U \iff SSE_R - SSE_U \geq 0$$

If H_0 is true then $SSE_R - SSE_U$ will be close to zero. A large $SSE_R - SSE_U$ is evidence that the null hypothesis isn't true

If the null hypothesis H_0 contains J restrictions (equations), then under H_0:

$$\boxed{\frac{(SSE_R - SSE_U)/J}{SSE_U/(n-K)} \sim F(J, n-K)} \rightarrow \text{can } F(J, n-K) \text{ be negative?}$$

where

\# of restrictions degrees of freedom

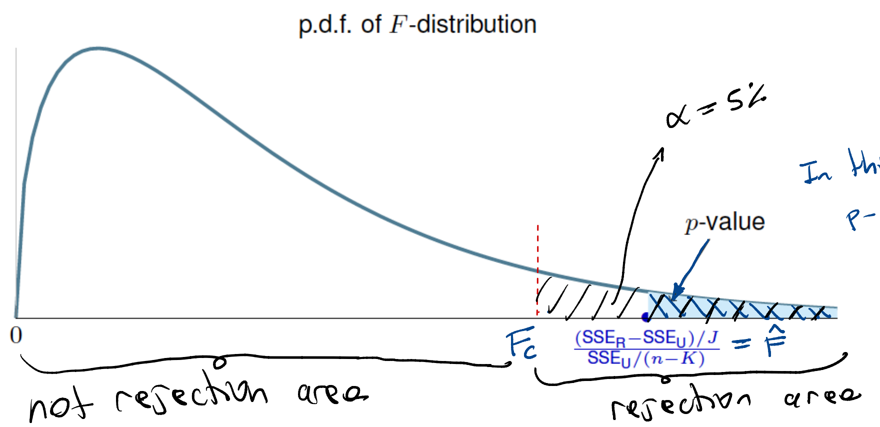
- $F(J, n-K)$ denotes the F distribution with J and $n-K$ degrees of freedom;
- J is the number of restictions in the null (count the number of equalities in the null);
- n is the sample size;
- K is the number of parameters (β's) in your model

(Notation : # = number of)

(in some textbooks σ_ε is also included as a parameter) $\Rightarrow$ #β's + 1 → see your lecture notes from class and follow that convention

$\boxed{\text{Important!!}}$ (could be #β's or #β's + 1)

From what follows I will use #β's



In this case p-value < 5%

A cumulates probability in that direction, from right to left

Compute $\hat{F} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-K)}$ from the sample

Let F_c denote the critical value of $F(J, n-K)$ distribution

Reject null hypothesis if

- $\hat{F} > F_c$ (indicates the $\hat{F}$ computed under the null is too extreme to be likely);
- **equivalently**, compute the p-value

$$\boxed{p\text{-value} = P\left(F(J, n-K) \geq \hat{F}\right)}$$

(Intuition) what is the likelihood that the random variable $F(J, n-K)$ is greater or equal to $\hat{F}$ ⇒ p-value

  and reject the null if p is less than the given significance level α

Some observations:

- Dividing by SSE_U is similar in spirit to dividing by standard error in simple hypothesis tests
- In practice, the p-value of a joint hypothesis test is often very small → very easy to reject
- If we reject the null hypothesis, this does *not* mean that we believe every equality is false! One equality or more is likely false, but one (or more) may be true...we simply can not distinguish which hypothesis in the null is false

Some notable special cases of F-tests include

(a) The first example above:

$$H_0 : \beta_2 = 0, \beta_3 = 0, \cdots, \beta_K = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0 \text{ or } \cdots \text{ or } \beta_K \neq 0$$

This hypothesis test is automatically run and reported in R

(b) Testing a single restriction, such as $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$. In this case, the F-test is equivalent to the familiar t-test

F test for 1 restriction: $\hat{F} = (\hat{t})^2$ Same relationship for both, statistic and probability distribution

just for the 1 restriction case!

and $F(1, n-K) = t^2(n-K)$

Practice Questions

Exercise 1

Suppose we collect $n = 1000$ i.i.d. observations on grades and time spent studying, attending lecture, and attending TA sessions: $\{(grade_i, study_i, lecture_i, TA_i)\}_{i=1}^{n}$. We estimate the following model:

$$grade_i = \beta_1 + \beta_2 study_i + \beta_3 lecture_i + \beta_4 TA_i + e_i$$

In our output we get the following results:

$$b_1 = 30, \qquad b_2 = 4, \qquad b_3 = 2.5, \qquad b_4 = 2, \qquad \underbrace{\sum_i (grade_i - \widehat{grade}_i)^2}_{=SSE_u} = 334,884.26 \quad \text{(unconstrained)}$$

$$\widehat{Cov}(b_1, b_2, b_3, b_4) = \begin{pmatrix} \widehat{Var}(b_1) & \widehat{Cov}(b_1, b_2) & \widehat{Cov}(b_1, b_3) & \widehat{Cov}(b_1, b_4) \\ \widehat{Cov}(b_2, b_1) & \widehat{Var}(b_2) & \widehat{Cov}(b_2, b_3) & \widehat{Cov}(b_2, b_4) \\ \widehat{Cov}(b_3, b_1) & \widehat{Cov}(b_3, b_2) & \widehat{Var}(b_3) & \widehat{Cov}(b_3, b_4) \\ \widehat{Cov}(b_4, b_1) & \widehat{Cov}(b_4, b_2) & \widehat{Cov}(b_4, b_3) & \widehat{Var}(b_4) \end{pmatrix} = \begin{pmatrix} 10 & 0.2 & 0.4 & -0.2 \\ 0.2 & 1.30 & -1 & 3 \\ 0.4 & -1 & 2 & -0.4 \\ -0.2 & 3 & -0.4 & 3 \end{pmatrix}$$

A student never studies or attends TA sessions. He claims that both studying and TA sessions are useless, and that by attending 20 hours of lecture (with zero time studying and zero time in TA sessions) his expected grade is 90 .

(a) State the student's null hypothesis and the natural alternative hypothesis.

(b) We estimate the model subject to the null hypothesis and find $SSE_R = 345,868.86$. Provide an expression for the p-value of the test, including a complete description of the relevant distribution.

(c) The student got a B on the midterm and is distraught. He now says studying is useless, and now makes no claim about lecture or TA sessions. State the new null hypothesis against a two-sided alternative.

(d) We estimate a model enforcing the new null hypothesis (from (c)) and find $SSE_R =$ 347.298,87 [redacted]. Think about two methods of testing the hypothesis in part (c) and provide two corresponding expressions for the p-value.

(a) constructing H_0:

(i) Study does not affect the grade $\Rightarrow \beta_2 = 0$

(ii) TA sessions does not affect the grade $\Rightarrow \beta_4 = 0$

$$E[grade_i \mid study_i = 0, lecture_i = 20, TA_i = 0] = 90$$

$$\Longleftrightarrow \text{(iii)} \quad \beta_1 + 20 \cdot \beta_3 = 90$$

• Null hypothesis

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_4 = 0 \\ \beta_1 + 20 \cdot \beta_3 - 90 = 0 \end{cases}$$

• alternative hypothesis

$$H_A : \begin{cases} \beta_2 \neq 0, \text{ or} \\ \beta_4 \neq 0, \text{ or} \\ \beta_1 + 20\beta_3 - 90 \neq 0 \end{cases}$$

(b) $SSE_R = 345,868.86$

$J = 3$ restrictions

$df = n - K = 1,000 - 4 = 996$ (df = degrees of freedom)

$$\hat{F} = \frac{(SSE_R - SSE_u)/J}{SSE_u/(n-K)} = 10.89$$

Finding the P-value

$$\Pr(F_{3,996} > \hat{F}) = \Pr(F_{3,996} > 10.89) \approx 0.001$$

reject H_0

(c) $H_0 : \beta_2 = 0$

$H_A : \beta_2 \neq 0$

(i) Could use a t-test

under the null $\beta_2 = 0$

$$\hat{t} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\hat{V}(\hat{\beta}_2)}} = \frac{4 - 0}{\sqrt{1.3}} = 3.5082$$

(ii) Could use a F test

$$\hat{F} = \frac{(SSE_R - SSE_u)/J}{SSE_u/(n-K)}$$

New $SSE_R = ?$

need $\widehat{grade\_restricted}$, that is we need new estimates b_1^R, b_3^R, b_4^R

$b_2^R = 0$ (restriction)

also need values of $\{TA_i, Lecture_i\}_{i=1}^{n}$

(d) We are given New $SSE_R = 347{,}298.87$

$$\hat{F}_{new} = \frac{(New\ SSE_R - SSE_u)/J}{SSE_u/(n-K)} = 12.3078$$

$$\sqrt{\hat{F}_{new}} = 3.5082 = \hat{t} \text{ (from (c))}$$

Exercise 2

6 parameter

Consider the following wage equation

$$ln(wage) = \beta_1 + \beta_2 educ + \beta_3 educ^2 + \beta_4 exper + \beta_5 exper^2 + \beta_6 (educ * exper) + e$$

where the explanatory variables are years of education and years of experience. Estimation results for this variable are shown in Figure below. These results are from the 200 observations in the file cps5_small.

	Coefficient Estimates and (Standard Errors)				
Variable	**Eqn (A)**	**Eqn (B)**	**Eqn (C)**	**Eqn (D)**	**Eqn (E)**
C	0.403	1.483	1.812	2.674	1.256 ***
	(0.771)	(0.495)	(0.494)	(0.109)	(0.191)
EDUC	0.175	0.0657	0.0669		0.0997 ***
	(0.091)	(0.0692)	(0.0696)		(0.0117)
EDUC^2	−0.0012	0.0012	0.0010		
	(0.0027)	(0.0024)	(0.0024)		
EXPER	0.0496	0.0228		0.0314	0.0222 ***
	(0.0172)	(0.0091)		(0.0104)	(0.0090)
EXPER^2	−0.00038	−0.00032		−0.00060	−0.00031
	(0.00019)	(0.00019)		(0.00022)	(0.00019)
EXPER × *EDUC*	−0.001703				
	(0.000935)				
SSE	37.326	37.964	40.700	52.171	38.012
AIC	−1.619	−1.612	−1.562		−1.620
SC (BIC)		−1.529	−1.513	−1.264	−1.554

1. What restriction on the coefficients of Eqn (A) gives Eqn (B)? Use an F-test to test this restriction. Show how the same result can be obtained using a t-test.
2. What restrictions on the coefficients of Eqn (A) give Eqn (C)? Use an F-test to test these restrictions. What question would you be trying to answer by performing this test?
3. What restrictions on the coefficients of Eqn (B) give Eqn (D)? Use an F-test to test these restrictions. What question would you be trying to answer by performing this test?
4. What restrictions on the coefficients of Eqn (A) give Eqn (E)? Use an F-test to test these restrictions. What question would you be trying to answer by performing this test?
5. Based on your answers to parts (a)–(d), which model would you prefer? Why?

Solution

1) Compare coefficients:

Eqn A has interaction term (educ × exper), with coefficient −0.001703

Eqn B does not have this interaction term

Restriction: coefficient of (exper × educ) equals 0

$\Rightarrow \beta_6 = 0 \rightarrow$ 1 restriction

F-test

Unrestricted model: Eqn A

restricted model: Eqn B

$$\Rightarrow \hat{F} = \frac{(SSE_R - SSE_u)/J}{SSE_u/df} = \frac{(37.964 - 37.326)/1}{37.326/(200-6)} = 3.316$$

at 5%

$F_c(1, 194) = 3.8898$

Do not reject the null

(You could also use t-test for this question because $J = 1$)

at 5%

$t_c(194) = -1.9723$, $t_c^2(194) = F_c(1, 194)$

2) Comparing Eqn A to Eqn C

Absent in Eqn C:

- interaction (exper × educ)
- exper
- exper2

$\Rightarrow$ restrictions of the following fashion

$\beta_4 = 0$ (exper coeff term)

$\beta_5 = 0$ ($exper^2$ coeff term)

$\beta_6 = 0$ ((exper x educ) coeff term)

Eqn A : unrestricted

Eqn C : restricted

$SSE_u = 37.326$

$SSE_R = 40.700$

$$\hat{F} = \frac{(40.700 - 37.326) / 3}{37.326 / 194} = 5.8454$$

at 5%

$$F_c(3, 194) = 2.6512$$

$$\hat{F} > F_c(3, 194)$$

reject the null

3) Compare Eqn B and Eqn D

Absent in Eqn D:

- educ $\Rightarrow \beta_2 = 0$
- $educ^2 \Rightarrow \beta_3 = 0$

$\left.\begin{matrix} \beta_2 = 0 \\ \beta_3 = 0 \end{matrix}\right\} J = 2$

In this case Eqn B is the unrestricted $\Rightarrow$ need to take into account that the degrees of freedom may change

Eqn B: 5 parameters

$\Rightarrow df = 195$

$$\hat{F} = \frac{(52.171 - 37.964)/2}{37.964/195} = 36.2996$$

$\downarrow$

to large, reject H_0 for sure//

at 5%

$F_c(2, 195) = 3.0422$

4) Compare model A and E

Absent in E

- $educ^2 \Rightarrow \beta_3 = 0$
- $(exper \times educ) \Rightarrow \beta_6 = 0$

$\left.\begin{matrix} \\ \\ \end{matrix}\right\} J = 2$

- A: unrestricted model $(df = 194)$
- E: restricted model

$$\hat{F} = \frac{(38.012 - 37.326)/2}{37.326/194} = 0.1226$$

at 5%

$F_c(2, 194) = 3.0425$

Do not reject the null

5) Trade-off $\rightarrow$ how well the model fits the data (+)

$\searrow$ how complex the model is (−)

AIC tends to select larger models

BIC tends to select smaller models

In this exercise, both AIC and BIC favour model E, t-statistic for each parameter also favours E //