7 – Joint Hypothesis Tests

Now that we have multiple "*X*" variables, and multiple βs, our hypotheses might also involve more than one β.

- We shouldn't use *t*-tests
- We should use the *F*-test

The types of hypotheses we are now considering involve multiple coefficients (βs). For example:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_A : \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0$$

and

$$H_0 : \beta_1 = 1, \beta_2 = 2, \beta_4 = 5$$
$$H_A : \beta_1 \neq 1 \text{ and/or } \beta_2 \neq 2 \text{ and/or } \beta_4 \neq 5$$

Note that the null hypothesis is wrong if *any* of the individual hypotheses about the βs are wrong. In the latter example, if $\beta_2 \neq 2$, then the whole thing is wrong. Hence the use of the "and/or" operator in H_A. It is common to omit all the "and/or" and simply write "not H_0" for the alternative hypothesis.

- A joint hypothesis specifies a value (imposes a restriction) for two or more coefficients
- Use q to denote the number of restrictions ($q = 2$ for 1st example, $q = 3$ for second example)

F-tests can be used for *model selection*. Which variables should we leave out of the model?

- If variables are insignificant, we might want to drop them from the model
- Dropping a variable means we hypothesize its β is zero
- Dropping multiple variables at once means all of the associated βs are all zero

Example: CPS data again

```
summary(lm(wage ~ education + gender + age + experience))
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.9574     6.8350  -0.286    0.775
education      1.3073     1.1201   1.167    0.244
genderfemale  -2.3442     0.3889  -6.028 3.12e-09 ***
age           -0.3675     1.1195  -0.328    0.743
experience     0.4811     1.1205   0.429    0.668
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.458 on 529 degrees of freedom
Multiple R-squared:  0.2533, Adjusted R-squared:  0.2477
F-statistic: 44.86 on 4 and 529 DF,  p-value: < 2.2e-16
```

The results of the above regression make me want to drop `age` and `experience`.

This corresponds to the hypothesis:

$$H_0\text{: } \beta_3 = 0 \text{ and } \beta_4 = 0$$

$$H_A\text{: either } \beta_3 \neq 0 \text{ or } \beta_4 \neq 0 \text{ or both}$$

Why would we want to drop variables?

We can't use *t*-tests

Idea (doesn't work): reject H_0 if **either** $|t_3|$ > 1.96 **and/or** $|t_4|$ > 1.96.

Review: type I error

Exercise: Assuming that t_3 and t_4 are *independent*, show that the type I error for the above test is 9.75% (not 5%).

How would you correct this problem? (Bonferroni method – not used in practice)

A bigger problem: t_3 and t_4 are likely *not* independent

In the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

- suppose that X_3 and X_4 are *not* independent (e.g. they are correlated)
- then the OLS estimators b_3 and b_4 will be correlated - the formula for b_3 (etc.) involves *all* of the "X" variables (remember OVB)
- then t_3 and t_4 will be correlated!

Example

Suppose that X_3 and X_4 are positively correlated. Consider the null:

$$H_0\colon \beta_3 = 0 \text{ and } \beta_4 = 0$$

- if b_3 and b_4 are both positive (or negative), it's not that big of a deal
- if one is positive and the other negative, that's a big deal

CPS data again

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -1.9574     6.8350  -0.286    0.775    
education      1.3073     1.1201   1.167    0.244    
genderfemale  -2.3442     0.3889  -6.028 3.12e-09 ***
age           -0.3675     1.1195  -0.328    0.743    
experience     0.4811     1.1205   0.429    0.668    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- do the signs of the coefficients make sense?
- what is the sign of the correlation between `age` and `experience`?
- according to the two individual *t*-tests, we fail to reject the null:

$$H_0\text{: } \beta_3 = 0 \text{ and } \beta_4 = 0$$

Let's try the *F*-test

I'm going to estimate two models:

- One model under the **alternative** hypothesis – we'll call the **unrestricted model** (the βs are allowed to be anything)

- One model under the **null** hypothesis – called the **restricted model**. I get this model by taking the null hypothesis to heart. That is, substitute in the values $\beta_3 = 0$ and $\beta_4 = 0$ into the full model

Unrestricted model (under H_A):

```
unrestricted <- lm(wage ~ education + gender
    + age + experience)
```

Restricted model (under H_0):

```
restricted <- lm(wage ~ education + gender)
```

F-test command:

```
anova(unrestricted, restricted)
```

Output (F-stat in blue, *p*-val in red):

```
Analysis of Variance Table

Model 1: wage ~ education + gender + age + experience
Model 2: wage ~ education + gender
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    529 10511
2    531 11425 -2   -914.27 23.007 2.625e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation? (A big *F*-stat still means reject)

A formula for the F-test statistic

- The F-test takes into account the correlation between the estimators that are involved in the test
- Note that if the unrestricted model "fits" significantly better than the restricted model, we should reject the null.
- The difference in "fit" between the model under the null and the model under the alternative leads to a formulation of the F-test statistic, for testing joint hypotheses.

The RSS is a measure of fit:

$$RSS = \sum_{i=1}^{n} e_i^2$$

where

$$\mathrm{e}_i = Y_i - \hat{Y}_i$$

The F-test statistic may be written as:

$$F = \frac{(RSS_{restricted} - RSS_{unrestricted})/q}{RSS_{unrestricted}/(n - k_{unrestricted} - 1)}$$

where q = # of restrictions, k = # of “X”s

Notice that if the restrictions are true (if the null is true), $RSS_{restricted} - RSS_{unrestricted}$ will be small, and we'll fail to reject.

Another statistic which uses RSS is the R^2:

$$R^2 = 1 - \frac{RSS}{TSS}$$

This gives us another formula for the F-test statistic:

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted} - 1)}$$

where:

$R^2_{restricted}$ = the R^2 for the restricted regression

$R^2_{unrestricted}$ = the R^2 for the unrestricted regression

q = the number of restrictions under the null

$k_{\text{unrestricted}}$ = the number of regressors in the unrestricted regression.

The bigger the difference between the restricted and unrestricted R^2's – the greater the improvement in fit by adding the variables in question – the larger is the F statistic.

Testing you on the exam

- The *F*-test statistic can be obtained by comparing the R^2 in the restricted model (H_0 model) and the unrestricted model (H_A model).
- The decision to reject or not depends on whether the *F*-stat exceeds the (5%) critical value:

q	5% critical value
1	3.84
2	3.00
3	2.60
4	2.37
5	2.21

- These values are only accurate if *n* is large (we'll always assume this)

Exercise

Test

$$H_0: \beta_3 = 0 \text{ and } \beta_4 = 0$$

in the model:

$$wage = \beta_0 + \beta_1 education + \beta_2 genderfemale + \beta_3 age + \beta_4 experience + \epsilon$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -1.9574     6.8350  -0.286    0.775    
education      1.3073     1.1201   1.167    0.244    
genderfemale  -2.3442     0.3889  -6.028 3.12e-09 ***
age           -0.3675     1.1195  -0.328    0.743    
experience     0.4811     1.1205   0.429    0.668    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.458 on 529 degrees of freedom
Multiple R-squared:  0.2533, Adjusted R-squared:  0.2477 
F-statistic: 44.86 on 4 and 529 DF,  p-value: < 2.2e-16
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.21783    1.03632   0.210    0.834    
education     0.75128    0.07682   9.779  < 2e-16 ***
genderfemale -2.12406    0.40283  -5.273 1.96e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.639 on 531 degrees of freedom
Multiple R-squared:  0.1884, Adjusted R-squared:  0.1853 
F-statistic: 61.62 on 2 and 531 DF,  p-value: < 2.2e-16
```