Linear models

Gabriela Ciołek

Testing multiple linear restrictions: the F test

- We wish to test multiple hypotheses about underlying parameters $\beta_1, \cdots, \beta_k$. We want to test whether a set of independent variables has no partial effect on a dependent variable.
- Unresticted model with k independent variables

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

  Number of parameteres in unrestiricted model is $k+1$.
- The null hypothesis is stated as

$$H_0 : \beta_{k-q+1} = 0, \cdots, \beta_k = 0,$$

  which puts q restrictions on the model.
- $H_1 : H_0$ is not true.
- Restricted model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + u.$$

F test

- $$F := \frac{(SSR\_r - SSR\_ur)/q}{SSR\_ur/(n-k-1)},$$
  where SSR_r is the sum of squared residuals from the resticted model and SSR_ur is the sum of squared residuals from the unrestricted model.
- $$q = \quad \textit{numerator degrees of freedom} \quad = df\_r - df\_ur$$
- $$n-k-1 = \quad \textit{denominator degrees of freedom} \quad = df\_ur$$
- One can show that under H_0, F is distributed as an F random variable with $(q, n-k-1)$ degrees of freedom
- We will reject H_0 in favor of H_1 when F is 'sufficiently' large (how large it depends of chosen significance level). The critical value depends on q and $n-k-1$

F test

- Once c has been computed, we reject H_0 in favor of H_1 at the chosen significance level if
$$F > c.$$
- If H_0 is rejected, than we say that $x_{k-q+1}, \cdots, x_k$ are jointly statistically significant at the appropriate significance level.
- Remark:The test alone does not allow us to say which of the variables has a partial effect on y; they may all affect y or maybe just only one affects y. If H_0 is not rejected, then the variables are jointly insignificant

Computing p-values for F Tests

- For reporting the outcomes of F tests, p-values are especially useful. Since the F distribution depends on the numerator and denominator df, it is difficult to get a feel for how strong or weak the evidence is against the null hypothesis simply by looking at the value of the F statistic and one or two critical values. In the F testing context, the p-value is defined as

$$p - value \quad = \mathbb{P}(\mathcal{F} > F),$$

  where $\mathcal{F}$ is an F random variable with $(q, n-k-1)$ degrees of freedom, and F is actual value of the test statistic.
- The same interpretation as it did for t statistics: it is the probability of observing a value of the F at least as large as we did, given that the null hypothesis is true. A small p-value is evidence against H_0.

Binary observations

Qualitative factors often come in the form of binary information: a person is female or male; a person does or does not own a personal computer; a firm offers a certain kind of employee pension plan or it does not; a state administers capital punishment or it does not. In all of these examples, the relevant information can be captured by defining a binary variable or a zero-one variable. In econometrics, binary variables are most commonly called dummy variables, although this name is not especially descriptive.

Example

Consider the following simple model of hourly wage determination:

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + u.$$

We use δ_0 as the parameter on female in order to highlight the interpretation of the parameters multiplying dummy variables; later, we will use whatever notation is most convenient. In above model, only two observed factors affect wage: gender and education. Since $female = 1$ when the person is female, and $female = 0$ when the person is male, the parameter δ_0 has the following interpretation: δ_0 is the difference in hourly wage between females and males, given the same amount of education (and the same error term u). Thus, the coefficient δ_0 determines whether there is discrimination against women: if $\delta_0 < 0$, then, for the same level of other factors, women earn less than men on average.

Interactions

- Sometimes it is natural for the partial effect, elasticity, or semi-elasticity of the dependent variable with respect to an explanatory variable to depend on the magnitude of yet another explanatory variable.
- Consider

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i x_j + \cdots + \beta_j x_j + \cdots + \beta_k x_k$$

- the partial effect of x_i on y is

$$\frac{\delta(y)}{\delta(x_i} = \beta_i + \beta_j x_j$$

- If $\beta_j > 0$, then, there is an interaction effect between x_i and x_j.
- We want to test if $\beta_j = 0$.

Exercise 1

- Consider the model $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$.
- Test $H_0 : \beta_1 = \beta_2$
- Use $F-$test to test H_0.
- Compute SSR_{ur}
- Compute SSR_r Note that in order to compute SSR_r we consider the model $log(wage) = \beta_0 + \beta_1 tenure + u$.
- compute F-statistic
- note that th denominator degree of freedom is $n - 4$ and $q = 2$.
- Compute $p-$value
- Code : $k0 = 4$
  $p = fdis\_prb(F, 2, n - k0)$

Do we reject or accept H_0?

Exercise 2

- Test $H_0 : \beta_{educ} = 0$.
- Use $F-$test to decide if H_0 should be rejected or accepted.
- The unrestricted model is the same as in Excercise 1
- The restricted model $log(wage) = \beta_0 + \beta_1 exper + \beta_2 tenure + u$.
- Calculate SSR_r.
- Compute $F-$ statistic
- $k0 = 4, \; q = 1)$
  $p = fdis\_prb(F, 1, n - k0)$

Exercise 3

- *load wage1.raw*
- Consider the model

$$log(wage) = \beta_0 + \beta_1 female + \beta_2 married + \beta_3 educ + \beta_4 exper + \beta_5 tenure + u$$

- Test $\beta_{female} = 0$.

Exercise 4 -Interactions

- Estimate a model that allows for wage differences among four groups: married men, married women, single men, and single women.
- To do this, we must select a base group; choose single men.
- Define dummy variables for each of the remaining groups: marrmale, marrfem, and singfem (see next slide).

Create the variables

- $marmale = (1 - female) \times married$
- $marrfem = female \times married$
- $singfem = female \times (1 - married)$

Creation of variables

$$educ = X(:, 2);$$

$$exper = X(:, 3);$$

$$tenure = X(:, 4);$$

$$female = X(:, 5);$$

$$married = X(:, 6);$$

$$marrmale = (1 - female). * married;$$

$$marrfem = female. * married;$$

$$singfem = female. * (1 - married);$$

- Do a regression for the model

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure \\ + \beta_4 marrmale + \beta_5 marrfem + \beta_6 singfem + u$$

- Calculate SSR_{ur}
- Test if being a women has an effect on the wage
- Do a regression for restricted model:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 marrmale + u$$

- Compute SSR_r.
- Calculate F-statistic
- Is being a women has an effect on wage?

Exercise 5

- Add an interaction term between *female* and *education* to the model

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female + \beta_5 femeduc + u$$

- $femeduc = female. * educ$
- This allows the education premumium depend on gender
- Test $\beta_{female} = \beta_{femeduc}$.