Notes 14: Continuation of Hypothesis Testing

ECO 231W - Undergraduate Econometrics

Prof. Carolina Caetano

1 F-test

Last class we discussed how to answer scientific questions of the type: "is $\beta_1 = a$?" Now we move on to a different kind of question. Instead of asking whether β_1 is zero, or anything else, we will ask questions that include information about more than one of the parameters in the model. Take one of our usual models

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 sections + u,$$

and suppose assumptions 1, 2 and 3 are satisfied. One question could be: does class affect grades? You should know how to test that. The question could, however, be different: are classes and sections perfect substitutes? What is the hypothesis you would like to test?

This is a different kind of question. It involves more than one parameter in the model. The protocol to test this hypothesis is the same as before. You choose the significance level, and then there is a rejection rule. For the rejection rule, the same way as before, we have to come up with a test statistic and a critical value. The statistic for this test is called an F-statistic, and it has the formula

What are those things? What are the $\hat{u}_i$?

this is easy, those are the residuals from the OLS regression. What are the $\hat{u}_{ir}$? Those are the residuals from an OLS regression on the restricted model. A restricted model is a model which incorporates the null hypothesis. So, in our case, it would be

So, you run this regression, and take the residuals. What is q?

In our example there is only one equation: $\beta_1 = \beta_3$, but there could be more.

Notice: the F-statistic is never negative, so we don't need to get its absolute value. Why is the F-statistic never negative? You don't have to know this, but you should be able to figure it out if you want.

What is the critical value $c(\alpha)$? The critical value is the $(1-\alpha)\cdot 100$th percentile of the $F_{q,n-k-1}$ distribution. So, the rejection rule is: At the α significance level, reject if

So, we spoke about how to incorporate 1 restriction. How do you incorporate more? For example, a typical question is: do class attendance, office hours and sections matter at all? How would you express this question mathematically into a hypothesis?

Observe how the alternative hypothesis is special. We are testing all three conditions at the same time, and thus the alternative is if any of them fails.

So, the rejection rule is exactly the same. The only special details in the statistic are q and $\sum_{i=1}^{n} \hat{u}_{ir}^2$. What is q? ------

How would you build the restricted model?

In an exam, I can ask any combination of questions which can yield composite null

hypotheses. You can get more examples if you look at Midterm 2 and the Final in 2010, and the Final in 2011 (find them in the Downloads page). However, in real life, the most important joint hypothesis refers to whether a set of coefficients is zero. This hypothesis is so important that we have special nomenclature when we reject this null hypothesis. We say that the coefficients are **jointly significant**. Just as with the t-test case, this name applies only to the null that the coefficients are equal to zero. If the null was anything else, we just say that we reject the null.

Observe that if you want to test things at the same time, you should test them together using an F test, you shouldn't simply run a series of t tests. Why?

The explanation is that if you run each test at a time, you have a 0.05 likelihood that you reject the null when it is true. But you do this many times, so it adds up in a non-trivial way. The result is that a test that is just a succession of t tests at the α significance level is not a test of α significance level. It has a higher level of significance. You are kind of messing around with things and loosing the parameters of comparison that you understand. How would you interpret a rejection if you don't even know the significance level you are using?. No. Use an F test to do it jointly, so that you will be sure to know exactly what your significance level is.

2 Comments on hypothesis tests

- Do not confuse statistical significance with economic importance. Something is significant when we rejected that it was equal to zero. This doesn't mean that the effect is important. For example, suppose that we found that the effect of class attendance on grade is of 0.1 points. However, we have so much data that our standard errors are really tiny, so we reject that the effect is zero. In other words, it seems as if the coefficient of classes is significant. Does that mean it is important? No. Even if each class does increase your grade 0.1 points, that doesn't matter very much, does it? If you go to all classes, say 20, then you will have gotten 2 more points in the final grade. That seems like a big waste of time for such a small return. In this case, it would be safe to say that classes are not really important for the grade.

- Do not ever, under any circumstance, decide the significance level after you look at the t-statistic. The significance level has to do with the level of certainty to which a science is accustomed. You should know this in advance. It is safe to say that in economics, though we still use 0.05 more than any other level, results that are

barely significant are considered inconclusive. In the end, we rely on results that have t-statistics of at least 3, hopefully more.

- A test of hypothesis operates under assumptions 1, 2 and 3 of the linear model. It is very hard to interpret your test results if the assumptions are wrong. With enough experience, you can develop the ability to use the test information even when the model is wrong, but only for certain kinds of departures from the model we studied. It is, in principle, possible for you to start making those kinds of analyses towards the end of this course, but it would require a lot of maturity with the material.

- There is a deeper understanding of a p-value. I am teaching it here, because it is better understood once you've been hearing about hypothesis tests for a while, and got used to the language. Remember that the p-value is the minimum significance level above which you would be rejecting the null. Suppose that $p = 0.02$. What is a significance level of 0.02? It's the probability of rejecting the null when it is true. It means that if the null was true, the probability of getting a t-statistic at least as large as what you got is 0.02. This is very, very low. It means that you would have to be very unlucky indeed. Probably it's not that. Probably the null was just wrong.

- That said, and this also relates to the first point in this section, don't forget that although the likelihood is small, it is not zero. A 10%, a 5%, even a 1% chance of anything bad is very unlikely, but it does happen, and it could happen to you. The only true protection against chance is to repeat studies. Does a different data set yield the same results? If you have a lot of observations, you could separate the sample in two, randomly. Do all your analysis in one sample, and get your conclusions. Now go and test it again in the other part of the sample. This strategy won't protect you against the faults of the data set itself, but it will give you a good measure of protection about chance flukes.

3 Summary

The order of scientific inquiry using the technology studied in this class is the following:

1. _____

2. _____. For this, I suggest a 4 step approach:

   (a) Begin by thinking of all the direct causes of the outcome. You don't need to know every cause exactly, but it helps if you know all the types of causes.

For example, birthweight is caused by genetics, health, substances consumed or to which the mother was exposed, and nutrition.

(b) Try to account for all the causes which are related to your variable of interest. If you observe them, then great. For example, in the birthweight example it is very common to observe drinking behavior, and several variables about the mother's health, including if she has diabetes, her weight and age.

(c) Try to approximate those factors that you cannot observe by using controls. For example: income, education and marital status and other variables are used to account for the mother's nutrition.

(d) Consider the linearity carefully. Should you use logs? Should you include the square of a control?

This looks very straight-forward, but it is not. Formulating a good model requires a lot of hard thinking, consulting the existing literature, studying the codebook of your data.

3. _____

_____.

Choose the significance level of your test. Depending on the type of test, make sure that you know what is the correct test statistic and critical value.

4. _____.

Also, estimate the standard errors. Finally, calculate the test statistic.

5. Compare test statistic to critical value to find out the answer of the test. Reflect on the implications of your test results for the economic problem your question is trying to address.

Notice that this is a rather naive procedure. Most researchers work with several models at the same time, test several hypotheses, and even report results for different significance levels, all in the same paper. Check the results table in the paper we are studying this year. It is very common to see several columns, each with a different model. Researchers report the estimates and the standard errors. Often they Put a * by the estimate to denote "significant at the 10% level," ** to denote "significant at the 5% level," and *** to denote "significant at the 1% level," or something similar to this. Also, researchers don't necessarily follow the procedure above in the exact order I suggested (though perhaps they should...).