Econometrics 1 (ECON 4003)
Suggested Solutions - Tutorial 7

Max Schroder

November 11, 2019

Abstract

This guide is supposed to be complementary to the official solutions supplied by the lecturer. All errors are my own.

Question 1

This question deals with hypothesis testing. You might have come across this before, but maybe now you will be able to understand it a little better. Hypothesis testing is the process of trying to make some statement about the **true** values of the regression parameters (β) using some information about our estimators ($\hat{\beta}$).

a)

Just to clarify the question here: we are interested in knowing whether there is a good chance that the **true** return to education is $\leq 8\%$, **given** that we have estimated a coefficient of 9%. See we have already learned that the value that we estimate will not necessarily be the true value of the parameter. However using the t-distribution we have a way of quantifying exactly how close we are likely to be.

This question uses the notion of "t-statistic" in two slightly different ways: 1. The "t-statistic under the null hypothesis" i.e.

$$t = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \tag{1}$$

2. the "t-statistic for the slope coefficient":

$$\frac{\hat{\beta}}{SE(\hat{\beta})} \tag{2}$$

this is a specific statistic that is sometimes reported by statistical packages (e.g. there's a reporting option for it in STATA).[1] You can think of it as the t-statistic under the null hypothesis that the true $\beta = 0$.

[1] Be careful not to confuse those with standard errors.

b)

There is always a little bit of controversy on the use of one sided hypothesis tests. On the one hand there can be important theoretical reasons for thinking that we can only err in one direction. On the other hand, one sided tests are often seen as suspicous since they make it a lot easier to reject a given null hypothesis. To illustrate this have a look at the picture below:

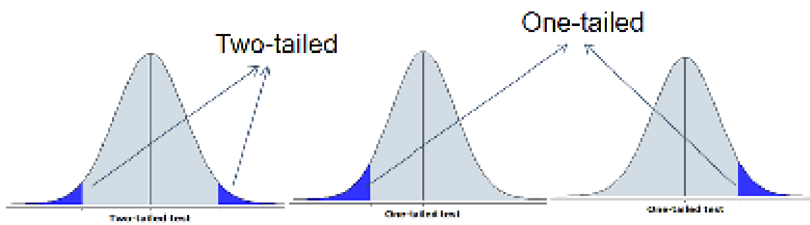

Figure 1: Source: https://analystprep.com/

Rejecting a null hypothesis effectively means that we obtain a t-value that falls somewhere in the blue-shaded area. As you can see there is a lot less area on either side of the distribution for a two sided test, making it harder to "hit" that region and consequentely to reject the null.

c)

A confidence intervall (CI) is in some sense the converse of a rejection region. You can think about it in this way. Ask yourself: "how far could I push β away from $\hat{\beta}$ without rejecting the null hypothesis that $\beta = \hat{\beta}$?"

Question 2

a)

OLS predictions give you the mean, as for any given vector of observables our best guess for the value of u is $E(u) = 0$. There are many applications however, where the population mean effect is not particularly interesting. If you want to play around a little I suggest reading the documentation of the "*margins*" command in STATA.[2]

b)

This might actually be one of the applications where a one sided hypothesis test would make sense. After all, we do not think that the gender pay gap is defined

[2] Or for the more ambitiously minded try a google search for "quantile regressions". This might provide inspiration for anyone still looking for a suitable disseration topic.

by men earning less than women.

c)

Maybe to add some mathematical rigour to the intuitive explanation of CIs above. Let x be a potential value of β and we are interested in testing the hypothesis that $\hat{\beta} = x$ at the 95% level. then we reject H_0 if either:

$$\frac{\hat{\beta} - x}{SE(\hat{\beta})} > t_{crit,0.025} \tag{3}$$

or

$$\frac{\hat{\beta} - x}{SE(\hat{\beta})} < -t_{crit,0.025} \tag{4}$$

if you solve these two equations for the values of x that exactly satisfy these inequalities, you get the CI formula.

Question 3

Just a general point on hypothesis tests and multiple comparisions, even though it is not strictly relevant for this question. Remember what it means to reject a null hypothesis at the $X\%$ level. It means that we are $1 - X\%$ certain that the result that we are seeing has not come about purely by randomness. This however leaves a (sometimes not so insignificant) margin open for chance. So if we have run a regression and test 20 different hypotheses at the 95% level we are likely to reject one of them simply due to random chance. This has been bothering statisticians for a long time and they have made some headway in addressing these issues, by adjusting the criticla values you'll have to find to reject the null (e.g. Bonferroni correction).

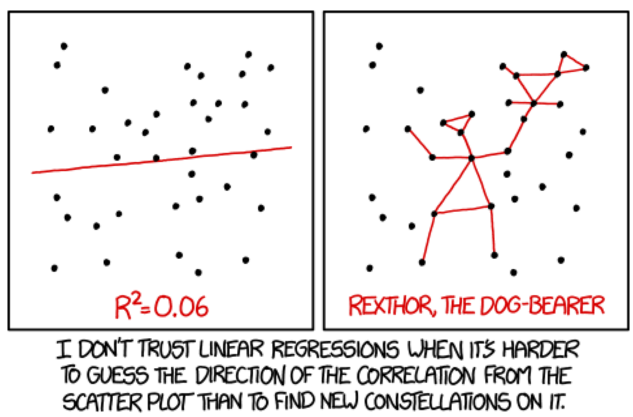

Figure 2: Source: https://xkcd.com/