

STATISTICAL GRAMMAR

A primer on interpreting regression coefficients without accidentally lying or spouting nonsense

By Sarah Hamersma
first draft 5/28/18; updated 7/6/2022

There are a lot of rules we learn in econometrics for policy evaluation. Things like:

- 1) If the coefficient is more than $1.96 \times$ the standard error, the estimate is statistically significantly different from zero.
- 2) If the 95% confidence interval contains zero, the estimate is not statistically significantly different from zero.

...and so on.

Many courses test your ability to apply these rules, and it's important to know them.

There are also rules of writing up results – a grammar of interpretation – that is just as important as the analysis itself. The four items below are some key points I'd like to emphasize as we think about interpreting results when running a regression of some variable Y on some covariate X , i.e. a model like $Y = \alpha + \beta X + e$ (though these rules apply across a variety of methods).

1. **“Significant” is already a word in English**
2. **“Statistically significant” is an adjective that can only modify an estimate**
3. **“Statistically insignificant effects” are not a thing**
4. **We just don't know the true effect**

So let's begin!

1. **“Significant” is already a word in English**

The tests above allow you to see whether a coefficient is statistically significantly different from zero. Many times, this long phrase is abbreviated to “statistically significant” – which may be all right, as long as you and your reader are absolutely clear on what that means. Too often, this long phrase is abbreviated in an extreme way: the estimate is declared “significant.”

Here's the thing. The word “significant” already has a meaning. It means “important” or “momentous” or “compelling” – it is something “of consequence.” It is a bad idea to utilize this word for something else. As McCloskey has taught us (in *Economical Writing*): if your reader finds you to be unclear, you are. You should write so that you cannot be misunderstood.

The following illustrates the problem:

“The estimated effect of the training intervention on job placements was significant.”

Your reader has the right to think you mean there was a big, meaningful effect. If your effect is a 0.01 increase in job placements from a \$50,000/person program, with a standard error of 0.005, what you have is an estimated effect of job placements that is statistically significantly greater than zero, but is, at the end of the day, insignificant. (In fact, this is nearly what we would call a “precise zero” – the confidence interval is very narrow and includes only very small numbers. We are 95% confident that this little interval contains the true effect, so we have reason to think the true effect is very small.) This sentence can be fixed by inserting the word “statistical” to make clear what you mean; you should also discuss the magnitude of the results in more detail.

2. “Statistically significant” is an adjective that can only modify an estimate

There are many appropriate ways to use the phrase “statistically significant” when writing up results from a coefficient that is statistically significantly different from zero. These would be sentences like:

“The coefficient on job training is positive and statistically significant.”

“The estimate is positive and statistically significant.”

“The estimated effect of job training is positive and statistically significant.”

Those are pretty much the only ones we should see, but unfortunately, people like to make at least two other uses of the phrase “statistically significant” that are incorrect.

(a) Referring to theoretical value via hypothesis

I recently read a paper that noted the following in setting up their hypothesis:

“If policy X really improves performance, then β should be significantly positive.”

What is wrong here? Well, someone is mixing up the theory with statistics. There aren’t any statistics in the theory; the theory is related to the truth about the population. Our theory cannot help us predict the precision with which that effect will be estimated (which is where statistical significance comes from). Try: “If policy X really improves performance, as predicted by theory, then we expect that β is positive.”

(b) Referring to the true causal effect rather than the *estimate* of that effect

“The effect of job training is positive and statistically significant.”

This sounds fine, right? WRONG. The “effect” is the unknown thing we are trying to estimate. It is beta. We still don’t know what it is. We have an *estimated* effect; this estimated effect has statistical features, such as a standard error that can help us assess our level of confidence in the sign of the estimated effect. Saying that an “effect is statistically significant” is mixing up two different parts of speech: the effect is an unknown scalar, and statistical significance is a feature of statistical objects with distributions. It is like saying that purple is itchy. It is nonsense.

But surely, no one really NOTICES when we do such things – we *all* know what we *all* mean, right? I’m not so sure; read on to see if you agree.

3. “Statistically insignificant effects” are not a thing

I hope that the very title of this item made you shudder, since you realize that just as “statistically significant” cannot modify “effects,” surely “statistically insignificant” cannot either. But it is the frequent use of this language that makes me think we do NOT all know what we are all saying, or what we all mean. Let’s continue with our example of job training, but suppose that the coefficient on job training is 0.01 with a standard error of 0.02. Common sentences to describe such a finding include things like:

“Job training had an insignificant impact on job placement.”

“Job training had a positive but statistically insignificant impact on job placement.”

“Job training caused a 1% increase in job placement, but the effect was statistically insignificant.”

“Job training increased job placement by a statistically insignificant 1%.”

There are several serious problems this brings up:

- 1) Even though statistically insignificant estimates are, by definition, those for which we lack confidence in the sign, the last three of these treat the sign of the estimate as the sign of the underlying effect. The fact that we did the significance test and still wrote these sentences says we don’t understand what a significance test actually tests. That is a serious problem.
- 2) The words “effect” and “increase” and “impact” are all causal words that suggest we know something about what the policy is doing, when in fact we do not know whether the actual effect is positive, negative, or nothing.

- 3) The last two sentences actually interpret the estimate of 1% as if it were the (unknown) effect. This is dangerous even when the estimate IS statistically significant, since we still only have an estimate of the effect and not the real thing.

What needs to be done differently? Well, we might need to learn to write things like:

“We are not able to identify the sign of the effect of the program on job placement.”

So sad. Too sad for most. (And too unpublishable?) So we try things like:

“The estimate is statistically insignificant, but it is positive as predicted by theory.”

Well...technically, “positive” is modifying “estimate,” so you are not making a false statement here. Similarly you could use the word “coefficient” instead of “estimate.” But is this tiptoeing around really the best way?

I am convinced that if we properly use confidence intervals, we will do a much better job of avoiding accidental forays into overstating our findings (because, after all, that is what is being done in these examples and we ought to admit it).

“Using our estimates, we cannot reject a zero effect, and our findings suggest that the range from -3% to +5% likely contains the true effect.”

Technically, there is still the issue that we should test against different nulls (such as -2%, +4%, or others) if we really want to say more about anything but zero (our null hypothesis). I leave further discussion to someone even more in the statistical grammar weeds than me.

4. We just don't know the true effect and that's okay

Perhaps the most painful realization for someone trying to write accurately about statistical analysis is that no matter what we do, we will *not know* if we have recovered the true beta. In fact, we probably will not have recovered it (it's basically a probability zero event, right?). The thing we have will always be *betahat*. Now, *betahat* is very useful and can give us a lot of insight, but many times we pretend that it is beta. Suppose, for the sake of argument, that *betahat* is 0.05 with a standard error of 0.02. Likely write-ups might be things like:

“The job training program increased job placements by 5%.”

“There was a statistically-significant 5% effect of the training program on job placements.”

Unfortunately – along with the misplaced “statistically-significant” modifier of the word “effect” in the second one – we are treating our estimate as if we have recovered beta. Actually, though, we have a confidence interval of roughly .01 to .09 (i.e. 1% to 9%). Try these:

“We emerge confident that the job training program increased placements, and our estimates suggest the improvement was likely between 1% and 9%.”

“The estimated effect of the job training program on placements is about 5%, with a confidence interval of 1% to 9%.”

The key here is to use the result of the statistical significance test to express your confidence about *the sign*, because that is what the test is testing.

It is not unusual for us to follow a rule of thumb where we do the significance test and, if the estimate is statistically significant, we feel free to talk about the estimate. (I tell my students this, actually, to try to stop them from talking about the estimate when it's NOT statistically significant. Also, to try to get them to notice the magnitude of the coefficient.) But we need to be more careful – it is not a test of whether your estimate is right. It is a test of whether your estimate is likely recovering the correct *sign*. This is because we default to a null hypothesis of zero; if you want to (statistically) rule out other possible values of beta besides zero, you should do the test again with a different choice of null.

So let's review.

1. **“Significant” is already a word in English**
2. **“Statistically significant” is an adjective that can only modify an estimate**
3. **“Statistically insignificant effects” are not a thing**
4. **We just don't know the true effect and that's okay**

It is surprisingly difficult to learn to follow these rules when we write up our results. First, we always want to have meaningful results, and whether intentionally or unintentionally, we push hard – too hard – quite frequently. Second, writing within these constraints often requires a few more words, and it takes practice to make it read smoothly. Finally – and I think this is most important – we have to unlearn what we read day in and day out in our journal articles. While I try to teach these kinds of principles in classes, the fact is that my students learn to write from the articles they read. We need to collectively push ourselves harder to teach them well, not just in classes, but by example.

* please give me feedback! sehamers@maxwell.syr.edu