

Michel Bierlaire

EPFL

Differences from classical hypothesis testing



Hello, in this video, we will investigate the difference between classical hypothesis testing, that you have probably learned at your bachelor statistical course, and hypothesis testing for specification tests, when we are doing modeling. There are subtle differences that I would like to discuss in this video.

Notes

Summary



0m 04s

Classical hypothesis testing: example

Null hypothesis (H_0)

A simple hypothesis contradicting a theoretical assumption.

Lady testing tea

- ▶ Theory: a lady is able to tell if the milk has been poured before of after the tea in a cup.
- ▶ H_0 : the outcome of the taste is purely random.



Classical hypothesis testing amounts to pose a null hypothesis which is simple, and which is supposed to contradict a theoretical assumption. An example that was used by Fisher, the statistician who introduced hypothesis testing, is a context of Lady testing cups of tea. In this case the theory is that a lady is able to tell if the milk has been poured before or after the tea in a cup. This is what she claimed. This is the theoretical assumption. Now the simple hypothesis contradicting this simple assumption, is that the outcome of the tasting is purely random. So whenever she says that the milk has been poured before or after, is the result of a purely random decision. This is H_0 .

Notes

Summary



0m 27s

Specification testing: example

Null hypothesis (H_0)

A simple hypothesis contradicting a theoretical assumption.

Explanatory variable



- Theory: a variable explains the choice behavior.
- H_0 : the coefficient of the variable is zero.

In the context of specification testing, we also have the fact that the null hypothesis is a simple hypothesis contradicting a theoretical assumption. But in this case we are testing if an explanatory variable explains the choice behavior or not. So to make the parallel with the lady testing example from classical hypothesis testing, here, our theoretical assumption is that a variable explains the choice behavior. Therefore, the null hypothesis that you would like to test is contradicting this theory. H_0 is that the coefficient of the variable is zero. So that the variable does not explain the behavior.

Notes

Summary



1m 19s

Errors in hypothesis testing

Type I error

- ▶ H_0 rejected and H_0 true.
- ▶ Include an irrelevant variable.
- ▶ Loss of efficiency.

Type II error

- ▶ H_0 accepted and H_0 false.
- ▶ Omit a relevant variable.

When you do hypothesis testing, you always make errors. And these errors are of two types. They are usually called type I error, and type II error. A type I error occurs when you reject the null hypothesis, but it happens to be actually true. Symmetrically, a type II error is when you decide to accept the null hypothesis when in reality this hypothesis is false. What does it mean in the context of specification testing? The type I error, when you reject the hypothesis, means that you will include an irrelevant variable. So you will keep in the model a variable which in reality does not explain anything about the behavior. For type II error, because you accept the null hypothesis, it means that you accept the fact that the variable does not play any role, you remove the variable from the model. But because H_0 is false, this variable, in reality, plays a role and explains the behavior. So the outcome of this is that you will omit an important variable in the model. What is the impact of this? A type I error, if you include a variable which is irrelevant, Well basically you will lose statistical efficiency. It means that the estimate of your parameters will be less precise.

Notes

Summary



Errors in hypothesis testing

Type I error

- ▶ H_0 rejected and H_0 true.
- ▶ Include an irrelevant variable.
- ▶ **Loss of efficiency.**
- ▶ Cost: C_I .

Type II error

- ▶ H_0 accepted and H_0 false.
- ▶ Omit a relevant variable.
- ▶ **Specification error.**
- ▶ Cost: $C_{II} \gg C_I$.

Note

In classical hypothesis testing, $C_I \approx C_{II}$

But type II error, because you omit a relevant variable, then you have a specification error. And actually this happens to be more severe, because most of the theory, the statistical theory in econometrics assumes that the specification of the model is right. So that you can make some statistical tests and statistical analyses on the estimated parameters. So let's look at the impact of the error in terms of cost, let's say an abstract cost that would represent the impact of the error. Let's call C_I the cost or the impact of a type I error, and let called C_{II} the cost of a type II error. Actually, this cost is way higher than the cost of a type I error. So it's much more severe to do a specification error, then to include an irrelevant variable and to lose efficiency. And this is one of the major difference with classical hypothesis testing. Because, usually, in classical hypothesis testing, both types of error have almost the same impact.

Notes

Summary



3m 26s

Impact of an error

Probability of an error

$$P(\text{Type I}) = P(H_0 \text{ rejected} | H_0 \text{ true}) \quad P(H_0 \text{ true})$$

$$P(\text{Type II}) = P(H_0 \text{ accepted} | H_0 \text{ false}) \quad P(H_0 \text{ false})$$

Expected cost

$$\text{Expected cost} = P(\text{Type I}) \quad C_I \quad + \quad P(\text{Type II}) \quad C_{II}$$

So let's calculate the impact of an error. First let's calculate the probability of an error. The probability to make a type I error is equal to the probability that we reject the null hypothesis, given that the hypothesis is true, times the probability that H_0 is true. In hypothesis testing, it's common to denote the probability that H_0 is rejected given that H_0 is true using the notation "alpha". And let's call lambda the probability that H_0 is true. Now we can do the same for the probability of a type II error. The probability that H_0 is accepted and H_0 is false is equal to the probability that H_0 is accepted, given that H_0 is false, times the probability that H_0 is false. The first probability is usually denoted by beta in hypothesis testing, and the probability that H_0 is false is one minus lambda. So given that we have these probabilities of errors, we can now calculate the expected cost of making an error. This expected cost is the sum of the expected cost of making a type I error, which is the probability that it occurs times the cost of making a type I error. that we denoted by C_I , plus the same for type II error. So the probability of type II times the cost of a type II error.

Notes

Summary



4m 30s

Impact of an error

Probability of an error

$$P(\text{Type I}) = P(H_0 \text{ rejected} | H_0 \text{ true}) \quad P(H_0 \text{ true})$$

$$P(\text{Type II}) = P(H_0 \text{ accepted} | H_0 \text{ false}) \quad P(H_0 \text{ false})$$

Expected cost

$$\begin{aligned} \text{Expected cost} &= P(\text{Type I}) C_I + P(\text{Type II}) C_{II} \\ &= \alpha \lambda C_I + \beta (1 - \lambda) C_{II} \end{aligned}$$

Specification testing

$\lambda \approx 0.5$, $C_{II} \gg C_I$: larger α can be used.

Now we use the quantities that we have calculated above. So the probability of type I error is $\alpha \lambda$ times C_I , and the probability of a type II error is β times one minus λ . And now we have C_{II} . In classical hypothesis testing, λ is close to 1, so usually H_0 is the hypothesis that we really believe in. And as I said before, the cost or the impact of making a type I or type II error is basically the same. So it means that to minimize the expected cost, it's custom to take α which is low. In practice, it's usually taken to be 5 percents or 1 percent. We prefer small α . In specification testing, the situation is slightly different. First, we have no strong a priori knowledge about the correctness of the null hypothesis. We are not sure if this variable is actually playing an important role or not in the behavior. So let's say that in this case λ is 50/50. It may be true, it may not be true, we don't know. But as we discussed before, in this context the cost of a type II error is way larger than the cost of a type I error. Therefore, it's recommended to use a larger α in this context. So using α equals 5 percents, like it is commonly used for hypothesis testing, maybe too strong. And we want to use a larger α to avoid the fact that the type II error may have a big influence on the cost or the impact on the modeling scheme.

Notes

Summary



6m 01s



So now, I recommend you to do some practical tests in order to remember, and to get more familiar with the concept of hypothesis testing. And then, what we will do, is that we will go more specifically into the testing and into the tests that we will apply for discrete choice models.

Notes

Summary



7m 40s