Print

Outcomes of Statistical Tests

Now that you've learned about the role of sample size in interval estimation, we turn our attention to hypothesis testing. Hypothesis testing requires a decision as to whether the data provide evidence against the null hypothesis.

Because we're making decisions about population parameters based on sample estimates, this decision-making process involves some risk.

When you make a decision based on the results of a hypothesis test, there are four possible outcomes. Your decision can result in a false positive, true negative, false negative, or true positive. You might find that this language is a little strange, and you're not alone.

The words "true" and "false" relate to whether you made a good decision or committed an error.

The words "positive" and "negative" relate to whether the null hypothesis was rejected or not rejected. Let's explore these four outcomes, one at a time.

For context, consider a scenario in which a court is ruling on a defendant's guilt. The null hypothesis is that the defendant is innocent. For the first two outcomes, the null hypothesis is true. The defendant did not commit the crime. He (or she) is truly innocent.

When you commit a false positive, the null hypothesis is true, but you erroneously reject it.

The court rules that the defendant is guilty, but he is actually innocent. This is called a type I error. The probability of committing a type I error is α. This is the significance level of your test. You choose the significance level to control the type I error rate.

If you choose a significance level of 0.1 or 0.15, you are more likely to commit a false positive than if your significance level is set at a lower value. This is why the significance level is often set at 0.05 or less.

A true negative is when the null hypothesis is true and you do not reject it. You make the correct decision. For our example, the court rules that the defendant is innocent, and he actually is.

The probability of a true negative is $1-\alpha$, where again α is the significance level. For the next two outcomes, the null hypothesis is false. The defendant is guilty. He actually did commit the crime. In a false negative, the null hypothesis is false, but you fail to reject it.

The court finds the defendant to be innocent, but he is actually guilty. This is called a type II error. The probability of a type II error is denoted by β.

For a true positive the null hypothesis is false, and it is correctly rejected as false. You make the correct decision. The court finds the defendant is guilty as charged. The true positive rate is also known as the test's power.

The probability of a true positive, or power, is $1- \beta$. The probability of a true negative and the probability of a false positive sum to 1. Likewise, the probability of a true positive and the probability of a false negative sum to 1.

In hypothesis testing, we choose α and β to balance the risks of false positives and false negatives.

To control the false positive (or type I error) rate, α is often set to 0.05. You also want a high probability of rejecting a false null hypothesis. So power, or $1 - \beta$, is often set at 0.8 or 0.9. In some applications, when minimizing the type II error rate is critical, power is set to 0.95.

Consider a scenario in which your null hypothesis is that the process is on target at some specified value. If you commit a false positive, a type I error, you would conclude that the process is off target when it's actually on target.

This might lead to a lot of good product being falsely rejected, resulting in costs associated with unnecessary rework or scrap. As a result, type I errors are often referred to as the supplier's risk. The other possible error is a false negative, or a type II error.

In this scenario, the process is off target, but you fail to detect it.

You have missed the signal that the process is off target. You risk accepting bad product and sending it to the customer. This can result in substantial costs from returns, lost customers, and a loss of reputation.

Because of this, type II errors are also referred to as the consumer's risk. In this scenario, you might be more willing to commit a false positive than take a chance that the customer will receive bad product.

So you might set the significance level, alpha, at a higher value, such as 0.10 instead of 0.05. Let's consider a final scenario. Athletes are regularly tested for banned substances. A false positive could lead to serious consequences, such as suspensions and other penalties.

As a result, drug tests are often designed to result in more false negatives than false positives. The rationale is that it's better to miss a signal that an athlete uses the drug than to wrongly accuse and potentially damage the athlete's career.

Close