Chapter 2

April 22, 2024

What is the definition of the positive predictive value?

The probability of a significant result + given that the H_1 is true.

What is the definition of a false-positive?

Coming to the conclusion that a true effect exists (ergo the H_1 is true) although in truth the H_0 is correct. Often also called a Type 1 error and denoted as α.

What is the definition of a false-negative?

Drawing the conclusion that the H_0 of no effect is true, when the H_1 is actually correct. Often also called a Type 2 error and denoted as β.

What is the definition of a true positive?

Correctly concluding that a true effect exists, when the H_1 holds true. Can be written as $1 - \alpha$ since it is the complement of a Type 1 error.

What is the definition of a true positive?

Correctly concluding that a true effect does not exist, when the H_0 holds true. Can be written as $1 - \beta$ since it is the complement of a Type 2 error.

If you perform 200 studies, where there is a 50% probability H0 is true, you have 80% power, and use a 5% Type 1 error rate, what is the most likely outcome of a study?

The most likely outcome is that we will find a true negative. Not the most promising outcome, but remember that we a priori assume both H_1 and H_0 to be equally plausible. In the next question we will see how to improve this.

How can you increase the positive predictive value in lines of research you decide to perform?

One possibility is to test hypothesis that are a priori likely to be true.

Why is it incorrect to think that "1 in 20 results in the published literature are Type 1 errors"?

To understand why this is incorrect one needs to revisit the positive predictive values. Only if the H_0 is true and ALL of those findings are published, only then I can expect 1 out of 20 findings to be false-positives

What is the problem with optional stopping

If done unsystematicly it can strongly increase the risk of Type-1 errors. However, as long as you use sequential analyses you can still do it.

How do multiple tests inflate the Type 1 error rate, and what can be done to correct for multiple comparisons?

It inflates like this: $1 - (1 - \alpha)^k$.

We can control this inflation by using statistical methods like the Bonferroni (α/k) or Holm correction.

What is the difference between a union-intersection testing approach, and an intersection-union testing approach, and under which testing approach is it important to correct for multiple comparisons to not inflate the Type 1 error rate?

In an union-intersection testing approach I make a claim when at-least-one test is significant. In these cases, I need to correct for multiple comparisons to control the error rate. Opposite stands the intersection-union testing approach, in which I only make a claim when ALL performed tests are statistically significant. I do not need to control for multiple comparisons.

In a replication study, what determines the probability that you will observe a significant effect?

The statistical power. TBC

Which approach to statistical inferences is the Neyman-Pearson approach part of, and what is the main goal of the Neyman-Pearson approach?

It is a frequentistic approach with the main goal of error control.

How should error rates (alpha and beta) in a statistical test be determined?

The answer is strongly context dependent. In some situations it might be more beneficial to reduce the risk of making false positive claims, while in another context reducing the Type 2 error rate is warranted. Note that both error rates can be reduced at the same time, e.g. by aiming for an α of .05 and a power of .95.