Chapter 1

April 22, 2024

What determines the shape of the p-value distribution?

The shape of a p-value distribution depends both on the presence of an real effect and on the statistical power aka the probability to detect a significant effect, if there is a true effect.

How does the shape of the p-value distribution change when there is a true effect and the sample size increases?

It becomes right-skewed. P-values below .01 become more likely.

What is Lindley's paradox?

Lindley's paradox describes the phenomena that when statistical power increases some p-values below .05 will become more likely, if there is no effect compared to if one exists.

How are p-values of continuous test statistics (e.g., t-tests) distributed when there is no true effect?

They are uniformly distributed, meaning that observing each p-value is equally likely. For example observing a p-value smaller than .05 is 5%

What is the correct definition of a p-value?

The probability of the observed data or more extreme data, given that the H_0 is true.

Why is it incorrect to think that a non-significant p-value means that the null hypothesis is true?

P-values are statements about the probability of *data* not about a theory

Why is it incorrect to think that a significant p-value means that the null hypothesis is false?

Type-1 errors exist.

Why is it incorrect to think that a significant p-value means that a practically important effect has been discovered?

First of all it could be a Type-1 error. It is also possible that we found a very little effect, only because of a study's large sample size and high statistical power. Although some benchmarks exist on how to interpret and found effect as "important", most of the time it depends on the situation.

Why is it incorrect to think that if you have observed a significant finding, the probability that you have made a Type 1 error (a false positive) is 5%?

Since p-values are a frequentistic tool they are bound to long run probabilities. Only in the long run - e.g. I do the same study 1e9 times - I can with confidence say, that 5% of them are Type-1 errors.

Why is it incorrect to think that $1 - p$ (e.g., $1 - 0.05 = 0.95$) is the probability that the effect will replicate when repeated?

It is in general not possible to determine a priori if a study will replicate or not. Too many other factors influence the outcome of a study, most of them outside the researchers control.

What are differences between the Fisherian and Neyman-Pearson approach to interpreting p-values?

The Fisherian approach defines a p-value as the descriptive continuous measure of how compatible the observed data and the null hypothesis (no effect) are. The higher the p-value is, the more likely the null hypothesis should be accepted. Since this approach only test the H_0 against any unspecified alternative hypothesis, the p value can only be used to describe the probability of the observed or more extreme data under the null model. The Fisherian approach is also called "Significance Testing". One big limitation of it is the sole definition of a Null, not also of an alternative hypothesis.

What does the null model, or the null hypothesis, represent in a null-hypothesis significance test?

The default model that suggests that there is no effect/relationship between the variables being studies

We can not use a null hypothesis significance test to conclude there is no (meaningful) effect. What are some statistical approaches we can use to examine if there is no (meaningful) effect?

Setting a SEOI, using Bayes-factors and equivalence tests.