Sifting the evidence – what's wrong with significance tests?

Jonathan Lam

09/19/21

Contents

1 Notes

- **epidemiological**: relating to the branch of medicine which deals with the incidence, distribution, and control of diseases

- Positive outcomes are more likely to be reporeted than null results – one in twenty positives will be a false positive given the $\alpha = 0.05$ cutoff

- **We need to care more about the interpretation of the values than the arbitrary use of the 0.05 cutoff** – the number was an arbitrary standard value given by Fisher

- Fisher focused on **Type I error**: rejecting the null hypothesis when it was true (false negative)

- Neyman and Pearson also considered **Type II error**: accepting the null hypothesis when it is false (false positive)

  - The size of an experiment is determined by choosing enough samples to cause a small Type I and Type II error rate (which can be calculated)

- **"To use the Neyman-Pearson approach we must specify a precise alternative hypothesis"**

  - People tend to forget about this and only think about the p-value

- Scientists have tried to alleviate the stark division between significant and non-significant results by using confidence intervals, but these still use the 5% cutoff.

- **Power** is the compliment of Type II error: i.e., it is the probability of rejecting the null hypothesis when it is in fact false

- People may misunderstand that the p-value is the probability that the null hypothesis is true, but this is wrong due to the prior distribution of experiments that for which the hypothesis is correct (usually, 90% of the hypotheses are incorrect)

  - This is an example of Simpson's paradox

- We can often increase the power of studies by increasing the sample size or increasing the precision of the instruments

- When significance testing was invented, sample sizes were much smaller and the 0.05 threshold was more reasonable

- Subgroup analyses should be skeptically analyzed (due to misrepresentation/Simpson's paradox)g

- "In observational studies it should be rememberd that considerations of confounding and bias are at least as important as the issues discussed in this paper"

2 Journal club

- Publication bias is about sexiness

- People didn't like the subjectivity of having researchers interpret the results and not always use the 0.05 value

  - Neyman and Pearson didn't like that – they wanted to get rid of the subjectivity
  - Neyman and Pearson said that what is important is a specific hypothesis, then use an exact threshold and don't leave it up to subjective interpretation

- 90% of hypotheses are wrong – only if you're doing dumb experiments

- You can just pay some journals to publish your stuff

- "It is often perfectly possible to increase the power of studies by increasing either the sameple size or the precision of the measurements"

  – No – otherwise they would have already done these

- Difficulty of Bayesian approach is that numbers have to be given to priors – difficult