

Statistical Power and ANOVA - Recap

Introduction

You've covered quite a bit in this section and should be gearing up to start conducting your own hypothesis testing! Before moving on to that exciting realm, take a minute to review some of the key takeaways.

Key Takeaways

Remember that the section began where the last left off, examining the relationship between α , power, effect size, and sample size. As you saw, these 4 quantities form a deterministic relationship; know any 3, and you can calculate the fourth. While a lower alpha value will lead to fewer type I errors, and a higher power will lead to fewer type II errors, in practice these are often set to common default standards due to exploding sample sizes required to detect various effect sizes. Some common thresholds used are:

- Setting alpha equal to 0.05 (or 0.01)
- Requiring power values of 0.8 or greater

After a thorough investigation of this relationship, you then also saw an alternative t-test, Welch's t-test which can be used for comparing samples of different sizes or different variances. While the formula was a bit complicated, the most important piece to remember is that when the assumptions that sample size and sample variance are equal for the two samples is violated, use Welch's t-test rather than the Student's t-test.

Aside from ensuring that the assumptions of a t-test are met, it's also important to know how type I errors are compounded if you perform multiple tests. This is known as the multiple comparison problem and you saw that type I errors compound under multiple tests. So while the probability of a type I error is equal to α for any one test, the collective probability that there is at least 1 type I error continues to increase as you perform more tests, further detracting from the confidence that you have uncovered a meaningful relationship. In order to account for this, you can use stricter criteria when defining α such as the Bonferroni correction. Alternatively, ANOVA is equivalent to a 2-sided t-test when comparing two groups, but also generalizes appropriately to multiple group comparisons.

Summary

Remember that simply observing a low p-value is not meaningful in and of itself. There are a number of factors to take into consideration when interpreting the results of a statistical test, from alpha, power, sample size, effect size, and the formulation of the problem itself. Good hypothesis testing requires careful thought and design.