2002.3 Hypothesis Testing

Part of @Stats Index

Why though

Null Hypothesis

Construct a null hypothesis. Often 'no effect' for social science findings - e.g. a beta of 0.

Our goal is to figure out whether or not we can reject the null hypothesis H_0 in favor of alternative hypothesis H_A.

First, we arbitrarily decide α, a threshold for significance. Choosing $\alpha = 3\%$, for instance, means that, under 100 repeated samples, we would falsely reject the null about 3 times.

The literature uses 5% because Fisher did one time, and now it's a Nash equilibrium in the α-choosing game.

p-values

A confidence interval, but worse. The p-value is our estimate of: if the null hypothesis were true ('under the null'), how frequently would we observe results as extreme as ours, or more so?

For instance, imagine flipping a coin 10 times. You come out with 9 heads. Is this coin tilted towards heads? Of $2^{10} = 1024$ possible combinations, $\binom{10}{9} + \binom{10}{10} = 11$ have 9 or more heads. So if we were flipping a fair coin 10 times over and over again, we'd see 9 or more heads $11/1024 = 1\%$ of the time.

Test it

The testing is simple. Compare a test statistic - say a p-value - to a 'rejection region' based on α to see how likely it'd be to observe our data under the null (there are other test statistics, all basically the same). If it's sufficiently unlikely, reject.

Constructing the 'rejection region' is based on α, but also if a test is one- or two-sided. I can pick 3 different H_0:

- the coin is fair
- the coin is not biased heads
- the coin is not biased tails

Suppose we choose number 2. Then, the 'rejection region' is the area of the test-statistic distribution $[a, \infty]$ such that $F(a) = .95$. In the above coin example, that rejection region would be about 8 or more heads. So we can reject the null hypothesis. (We could also say our results are significant at the 1% level - whichever makes the finding look better I guess.)

If we had hypothesis 3, we could *not* reject the null. We have no evidence it is biased towards tails - the complement of H_0 (an example of needlessly discarding information - not only can we fail to reject H_0, but we can also say it's likely biased towards heads!).

Failing to reject the null is not the same as finding evidence for the null. It is evidence *against* whatever the alternative hypothesis was. It could be that H_A is the complement of H_0, but that's often not the case - consider the second and third H_A above.

Rejecting the null doesn't mean your finding is good. You could find a tiny effect with a great p-value, and nobody will care.

Difference in means test

$$\text{Test Statistic} = \frac{\mu_A - \mu_B}{\sqrt{S_A^2/n_A + S_B^2/n_B}}$$