Multiple comparisons

STA9750

Multiple comparisons

Suppose you have data coming from different groups

- Different medical treatments
- Different marketing campaigns
- Different instructors
- ...

ANOVA / Kruskal Wallis

Earlier in the semester, we saw ANOVA and we briefly mentioned Kruskal Wallis (NPAR1WAY)

ANOVA

- Assumes normality, equality of variances across groups
- Detects differences in means

Kruskal-Wallis

- Doesn't assume normality or equality of variances
- Detects differences in medians

Pairwise comparisons

ANOVA/Kruskal Wallis tells us if there are differences in the groups, but *don't tell us where*

Alternative: Do pairwise t-tests, and see if there are significant differences between the pairs

Problem:

- If there are no differences between the groups, the probability that any given test gives a false positive is 0.05 (assuming that's our significance level)
- If we do more than one test, the probability of having at least one false positive grows fast...
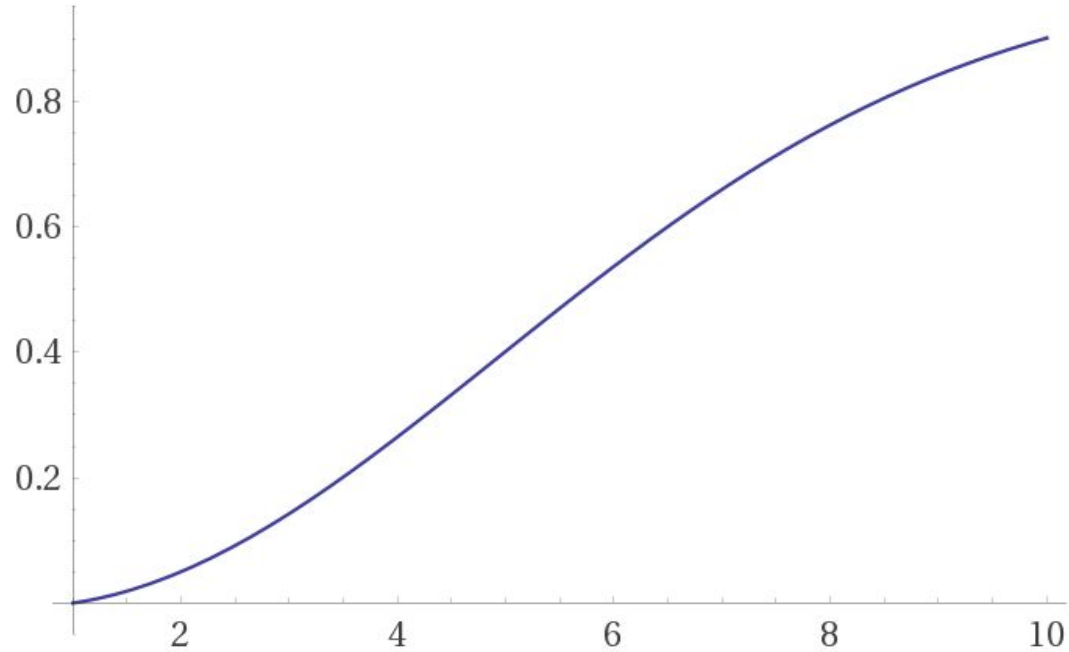
False positives grow fast

If we have k groups, there are k choose 2 pairwise comparisons

If our significance level is 0.05, the probability that there's at least one false positive is

Pr(at least one false positive) = 1 - Pr(no false positives)

$$= 1 - 0.95^{(k \text{ choose } 2)}$$

For example, if k = 5, Pr(at least one false positive) is approximately 0.4

Probability of at least one false positive

Computed by Wolfram|Alpha

Bonferroni

A general solution to this problem (which isn't specific for pairwise comparisons) is the following

Bonferroni: If you are doing N tests and you want to ensure an overall false positive rate of 0.05, conduct the tests at the 0.05/N significance level

Problem: Very stringent. For example, if we have 5 groups, there are N = 10 pairwise tests, so we should perform the tests at the 0.005 significance level, which is quite harsh

Tukey's Honest Significant Difference (HSD)

Specific procedure for normal means

Less stringent than Bonferroni

1) Find all pairwise differences in means
2) If the absolute difference between 2 means is greater than some threshold, they're significantly different
3) Otherwise, they're not

With Tukey, Pr(at least 1 false positive) = 0.05 (if that's our significance level)

There are other methods (Scheffe, Newman-Keuls, etc.). We won't cover them.

Tukey in SAS

In PROC ANOVA, add:

    means <variable> / tukey;

Multiple comparisons with a control

Suppose you have a control group and want to test whether a few treatments are significantly different than the control

You have multiple tests (one for each treatment, which can be different than the control or not), so you have to be careful to ensure that the probability of a false positive is controlled

Dunnett's test covers this situation

SAS command:

```
means <variable>  / dunnett('<control>');
```