Hypothesis Tests
Multiple Comparisons of the Mean

DASC 512

Overview

- When would this be used?
- Bonferroni
- Bonferroni-Holm
- Tukey’s HSD / Tukey-Kramer

Post-hoc Analysis

After conducting an ANOVA and rejecting the null hypothesis, we may be interested in doing pair-wise comparisons between groups.

We do this by generating confidence intervals on the difference between each pair of groups

For k treatments, there are $c = \frac{k(k-1)}{2}$ pairs of means to compare.

If we kept α steady for each comparison, our actual probability of a Type I error would balloon to $1 - (1 - \alpha)^c$

Bonferroni Approach

Bonferroni's approach is quick and easy, but may be too conservative.

Boole's inequality: For events $E_1, E_2, \ldots, E_c$ in a sample space, the probability of at least one event occurring has the following inequality.

$$P(E_1 \cup E_2 \cup \cdots \cup E_c) \leq P(E_1) + P(E_2) + \cdots + P(E_c)$$

Bonferroni Approach

$$P(E_1 \cup E_2 \cup \cdots \cup E_c) \leq P(E_1) + P(E_2) + \cdots + P(E_c)$$

Using this, the probability of at least one Type I error occurring is:

$$P(\text{Type I}) \leq \alpha + \alpha + \cdots + \alpha = c\alpha$$

Bonferroni's approach: Do t-tests with individual α set at $\frac{\alpha}{c}$

Bonferroni Approach

Recall our coughing treatment example. We had three treatments, so we have

$$c = \frac{k(k-1)}{2} = \frac{3(2)}{2} = 3$$

comparisons to conduct.

We can do two-sample t-tests or confidence intervals using $\alpha = \frac{0.05}{3} = 0.0167.$

Bonferroni-Holm

Holm updated Bonferroni to make it perform slightly better.

This method tests pairs sequentially starting at the Bonferroni α, then increases α for each consecutive pair until the last pair is tested at the original α

DM-Honey: $\alpha = \frac{0.05}{3}$

DM-Placebo: $\alpha = \frac{0.05}{2}$

Honey-Placebo: $\alpha = \frac{0.05}{1}$

Tukey's HSD

"Honest Significant Difference" is more advanced (and complicated) but generally gives tighter intervals than Bonferroni.

- Tukey only works for equal n_i
- Python implements Tukey-Kramer test when n_i is not equal

```
sm.stats.multicomp.MultiComparison(data,groups).tukeyhsd(alpha).summary()
```

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
====================================================
group1  group2 meandiff p-adj   lower   upper  reject
----------------------------------------------------
    dm   honey    1.981 0.0279  0.1756  3.7863   True
    dm placebo  -1.8198 0.0441 -3.6014 -0.0383   True
 honey placebo  -3.8008  0.001 -5.5552 -2.0464   True
----------------------------------------------------
```

Which one is best?

There's no accepted answer, but you should always pick your analysis method prior to collecting data.

Bonferroni or Holm are useful for a broader set of circumstances. I'd use Holm for any time you're conducting several hypothesis tests.

Tukey's HSD and Tukey-Kramer are designed for post-hoc ANOVA analysis. I'd use Tukey in those situations.

Recap

- When would this be used?
- Bonferroni
- Bonferroni-Holm
- Tukey’s HSD / Tukey-Kramer