Tukey's HSD and Bonferroni

When to use post-hoc tests

After ANOVA rejects H_0, we know *at least one* mean differs—but not *which* ones. Post-hoc tests make pairwise comparisons while controlling the overall Type I error rate.

Tukey's HSD (Honestly Significant Difference)

Use for all pairwise comparisons. Two means are significantly different if:

$$|\bar{x}_i - \bar{x}_j| > \text{HSD} = q_{\alpha,k,N-k}\sqrt{\frac{\text{MSW}}{n}}$$

where q = critical value from Studentized Range table, k = number of groups, N = total observations, n = group size (assumes equal sizes), and MSW = mean square within from ANOVA.
Example: From an ANOVA comparing three teaching methods ($n = 5$ per group):
$\bar{x}_A = 78.8, \quad \bar{x}_B = 86, \quad \bar{x}_C = 91.2, \quad \text{MSW} = 6.97$
With $q_{0.05,3,12} = 3.77$:

$$\text{HSD} = 3.77\sqrt{\frac{6.97}{5}} = 3.77(1.18) = 4.45$$

Pairwise comparisons:

Comparison	$\lvert\bar{x}_i - \bar{x}_j\rvert$	$>$ HSD?	Conclusion
A vs B	$\lvert 78.8 - 86\rvert = 7.2$	$7.2 > 4.45$	Significant
A vs C	$\lvert 78.8 - 91.2\rvert = 12.4$	$12.4 > 4.45$	Significant
B vs C	$\lvert 86 - 91.2\rvert = 5.2$	$5.2 > 4.45$	Significant

All three methods differ significantly from each other.

Bonferroni correction

Use when making a smaller number of planned comparisons. Adjust the significance level:

$$\alpha^* = \frac{\alpha}{m}$$

where m = number of comparisons. Then perform individual t-tests, rejecting if $p < \alpha^*$.
Example: With 3 groups, there are $\binom{3}{2} = 3$ pairwise comparisons. At $\alpha = 0.05$:

$$\alpha^* = \frac{0.05}{3} = 0.0167$$

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning

OntarioTech UNIVERSITY

Each pairwise t-test must have $p < 0.0167$ to be declared significant.

Example: A researcher has 4 groups but only wants to compare Group 1 vs Group 2 and Group 1 vs Group 3 (two planned comparisons). At $\alpha = 0.05$:

$$\alpha^* = \frac{0.05}{2} = 0.025$$

Which method to use?

Tukey's HSD	Bonferroni
All pairwise comparisons	Few specific comparisons
Equal sample sizes preferred	Any sample sizes
More power with many comparisons	More power with few comparisons

Tip: With many comparisons, Bonferroni becomes very conservative (hard to reject). Tukey's HSD is generally preferred for all pairwise comparisons.

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning