# Post Hoc Comparisons

## Data

The following data set reflects a between-subjects design with one factor (with three levels). The data are presented in a format suitable for entry into statistical software.

| | Factor | Outcome |
|---|---|---|
| 1 | 1.00 | .00 |
| 2 | 1.00 | .00 |
| 3 | 1.00 | 3.00 |
| 4 | 1.00 | 5.00 |
| 5 | 2.00 | 4.00 |
| 6 | 2.00 | 7.00 |
| 7 | 2.00 | 4.00 |
| 8 | 2.00 | 9.00 |
| 9 | 3.00 | 9.00 |
| 10 | 3.00 | 6.00 |
| 11 | 3.00 | 4.00 |
| 12 | 3.00 | 9.00 |

## Computer Output

The following tables represent typical output from statistical software. Options, labels, and layout vary from program to program.

The table of descriptive statistics can be used to determine the inferential statistics.

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Level 1 | 4 | 2.000 | 2.449 | 1.225 |
| Level 2 | 4 | 6.000 | 2.449 | 1.225 |
| Level 3 | 4 | 7.000 | 2.449 | 1.225 |

| Source | SS | df | MS | F | p | Eta^2 |
|---|---|---|---|---|---|---|
| Between | 56.000 | 2 | 28.000 | 4.667 | 0.041 | 0.509 |
| Within | 54.000 | 9 | 6.000 | | | |
| Total | 110.000 | 11 | | | | |

The table of inferential statistics shows the key elements to be calculated.

| Factor | Factor | Mean Difference | SE Difference | HSD | p | Lower CI | Upper CI |
|---|---|---|---|---|---|---|---|
| Level1 | Level2 | -4.000 | 1.732 | -2.309 | 0.106 | -8.836 | 0.836 |
| Level1 | Level3 | -5.000 | 1.732 | -2.887 | 0.043 | -9.836 | -0.164 |
| Level2 | Level3 | -1.000 | 1.732 | -0.577 | 0.835 | -5.836 | 3.836 |

## Calculations

Mean Differences: Mean Differences (raw effects) are the differences between the means for all pairs of groups. Even though half of the possible pairwise comparisons are redundant, the mean differences will have the opposite signs because of subtraction order.

$$M_1 - M_2 = 2.000 - 6.000 = -4.000$$
$$M_1 - M_3 = 2.000 - 7.000 = -5.000$$
$$M_2 - M_3 = 6.000 - 7.000 = -1.000$$

Standard Error of the Difference: These standard errors are for the difference between the two group means in each comparison. The values are a function of the $MS_{WITHIN}$ (from the ANOVA) and the sample sizes. [In this case, because all groups are of the same size, the standard error for each comparison is the same.]

$$SE_{DIFF} = \sqrt{\left(\frac{MS_{WITHIN}}{n_{GROUP}}\right) + \left(\frac{MS_{WITHIN}}{n_{GROUP}}\right)} = \sqrt{\left(\frac{6.000}{4}\right) + \left(\frac{6.000}{4}\right)} = 1.732$$

Statistical Significance: The HSD statistic is a ratio of the mean difference to the standard error of the difference. There is one statistic for each of the comparisons.

Because the ANOVA has $df_{BETWEEN} = 2$ and $df_{WITHIN} = 9$, $HSD_{CRITICAL} = 2.792$

$$HSD_{1v2} = \frac{M_1 - M_2}{SE_{DIFF}} = \frac{-4.000}{1.732} = 2.309$$

Because $HSD < HSD_{CRITICAL}$, $p > .05$

$$HSD_{1v3} = \frac{M_1 - M_3}{SE_{DIFF}} = \frac{-5.000}{1.732} = 2.887$$

Because $HSD > HSD_{CRITICAL}$, $p < .05$

$$HSD_{2v3} = \frac{M_2 - M_3}{SE_{DIFF}} = \frac{-1.000}{1.732} = 0.577$$

Because $HSD < HSD_{CRITICAL}$, $p > .05$

Confidence Intervals: For HSD, calculate the confidence intervals around (centered on) each mean difference separately.

$$CI_{1v2} = (M_1 - M_2) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -4.000 \pm (2.792)(1.732) = [-8.836, 0.836]$$
$$CI_{1v3} = (M_1 - M_3) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -5.000 \pm (2.792)(1.732) = [-9.836, -0.164]$$
$$CI_{2v3} = (M_2 - M_3) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -1.000 \pm (2.792)(1.732) = [-5.836, 3.836]$$

## APA Style

Post hoc tests build on the ANOVA results and provide a more focused comparison among the groups and usually follows a presentation of the ANOVA (which already includes the descriptive information). The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the confidence intervals (and can be presented on its own).

Tukey's HSD tests showed that the first group scored statistically significantly different than the third group, $t(9) = -2.89$, $p = 043$. However, the other comparisons were not statistically significant ($ps > .05$).

A series of Tukey's HSD comparisons revealed that the first group ($n = 3$, $M = 2.00$, $SD = 2.45$) scored substantially lower Outcome scores than the third group ($n = 3$, $M = 7.00$, $SD = 2.45$), 95% CI [-9.84, -.16], $t(9) = -2.89$, $p = 043$. However, the other comparisons revealed effectively little to no difference between the other groups ($ps > .05$).

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.