

Stat 218 - Day 30
More on ANOVA: Multiple Comparisons

Recall that analysis of variance is a technique for comparing means among several groups. We test $H_0: \mu_1 = \mu_2 = \dots = \mu_I$ by constructing an ANOVA table and determine the test statistic as

$$F_s = \frac{MS(\text{between})}{MS(\text{within})} = \frac{SS(\text{between})/(I-1)}{SS(\text{within})/(n^*-I)}, \text{ where } SS(\text{within}) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^I (n_i - 1)s_i^2$$

and $SS(\text{between}) = \sum_{i=1}^I n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$.

Multiple comparisons:

When an ANOVA F-test leads to rejecting the null hypothesis that all groups have the same population mean, the natural question to ask is “*which* groups’ means differ?” We can not just do a bunch of two-sample *t*-tests, because the probability of type I error goes up for each test that we perform. We need a procedure that will look for differences between pairs of groups with a global type I error rate of whatever we set α to be.

There are several different such **multiple comparison** procedures. We’ll learn to implement the *Tukey* method using Minitab. This procedure gives a confidence interval for the difference in group population means for all pairs of groups. The confidence level applies to all of the intervals *simultaneously*. The intervals that do not include the value zero suggest a significant difference between the group population means for that pair of groups.

Example: Discrimination against handicapped (cont.)

Recall that we found a test statistic of $F_s = 2.86$ and $P\text{-value} = .030$, so we have reasonably strong evidence that the five groups’ population means are not all equal. We can now investigate *which* groups have population means that differ significantly.

(a) Use Minitab to reproduce the ANOVA table (Stat> ANOVA> One-way...).

(b) Report the sample mean ratings, in order from smallest to largest:

(c) To produce the Tukey intervals, start with Stat > ANOVA > One-way..., but select the “Comparisons” option and check the box for the Tukey procedure with a family error rate of 5%. Record all ten of the confidence intervals that Minitab produces:

mu_crutches – mu_amputee:

mu_hearing - mu_amputee:

mu_none – mu_amputee:

mu_wheelchair – mu_amputee:

mu_hearing - mu_crutches:

mu_none – mu_crutches:

mu_wheelchair – mu_crutches:

mu_none – mu_hearing:

mu_wheelchair – mu_hearing:

mu_wheelchair – mu_none:

(d) Which of these intervals do not include the value zero? What does that reveal about which pairs of handicap types have significantly different (at the $\alpha=.05$ level) mean qualification ratings? Explain.

We sometimes represent which groups differ significantly from which others by underlining the groups whose means do not differ significantly from other groups. First let us arrange the groups in order of their sample means:

	Hearing	Amputee	None	Wheelchair	Crutches
Sample mean qualification rating	40.50	44.29	49.00	53.43	59.21

(e) Underline the groups whose mean qualification ratings do not differ significantly from each other.

Example: Hot dogs (cont.)

Use Minitab's Tukey procedure to determine which hot dog types differ significantly, at the $\alpha=.01$ level, with regard to mean calorie amounts. Then repeat the analysis with sodium content as the response variable.

Example: Crash test dummies (cont.)

Use Minitab's Tukey procedure to determine which number of door categories differ significantly, at the $\alpha=.05$ level, with regard to the log of head injury measurements.