STAT 252 – Handout 12
More on ANOVA: Multiple Comparisons, Transformations

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

$$F = \frac{MS(between)}{MS(within)} = \frac{SS(between)/(k-1)}{SS(within)/(N-k)},$$

where:

$$SS(within) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{k} (n_i - 1) s_i^2$$

and:

$$SS(between) = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2.$$

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 cannot 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 $\alpha$ 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 12-1: Employment Discrimination (cont.)**
Recall that we found a test statistic of $F = 2.86$ and $p$-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:

\[ \mu_{\text{crutches}} - \mu_{\text{amputee}} \]
\[ \mu_{\text{hearing}} - \mu_{\text{amputee}} \]
\[ \mu_{\text{none}} - \mu_{\text{amputee}} \]
\[ \mu_{\text{wheelchair}} - \mu_{\text{amputee}} \]
\[ \mu_{\text{hearing}} - \mu_{\text{crutches}} \]
\[ \mu_{\text{none}} - \mu_{\text{crutches}} \]
\[ \mu_{\text{wheelchair}} - \mu_{\text{crutches}} \]
\[ \mu_{\text{none}} - \mu_{\text{hearing}} \]
\[ \mu_{\text{wheelchair}} - \mu_{\text{hearing}} \]
\[ \mu_{\text{wheelchair}} - \mu_{\text{none}} \]

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.
Example 12-2: 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 12-3: Crash Test Dummies
The Minitab worksheet crash.mtw contains data on automobile crash test results. Response variables are measurements of injury extent on head (c5), chest (c6), left leg (c7), and right leg (c8). Explanatory variables include whether the dummy was on the driver or passenger side (c9), protective devices in the car (c10), number of doors on the car (c11), year of make (c12), and size of car (c14). We will first investigate whether the head injury measurements appear to differ based on number of doors.

a) Construct visual displays of the head injury measurements by the number of doors. Describe the distributions, paying particular attention to the question of whether the head injury measurements appear to differ significantly among the three groups.

b) Does it appear that the technical conditions of the ANOVA procedure and F-test are satisfied with these data? Explain.

- A common remedy for non-normally distributed data is to take a transformation that achieves approximate normality.
  - With data that are skewed to the right, taking a log transformation (i.e., using a log scale) is often helpful.
c) Transform the head injury measurements by using logarithms (base 10). Re-examine the distributions of these measurements by the number of doors. Do the technical conditions appear to be satisfied now? Explain.

d) Construct the ANOVA table and interpret the results. Do the data provide strong evidence that head injury measurements differ based on the number of doors on the vehicle? Explain.

e) 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.

The following (incomplete) ANOVA table pertains to the log of head injury measurements, with the vehicle’s year of make as the explanatory variable:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td></td>
<td>0.2152</td>
<td>0.0538</td>
<td>0.291</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>333</td>
<td>14.35724</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>337</td>
<td>14.5724</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

f) Fill in the four missing entries of the ANOVA table.

g) How many groups were there for year of make? Explain how you know.

h) What conclusion would you draw from this analysis? Explain.