Stat 302 – Day 13
Post-hoc Comparisons

Last Time: Analysis of Variance (“ANOVA”)
- Choice of test statistic: F-ratio = within group variability / between group variability
  With equal sample sizes: \( n_i \, \text{var}(\bar{x}_i) / \text{mean}(s^2) \)
  
  \[
  \frac{\sum_{i=1}^{I} n_i (\bar{x}_i - \bar{x})^2 / (I - 1)}{\sum_{i=1}^{I} (n_i - 1) s_i^2 / (n - I)} = \frac{\text{MSGroup}}{\text{MSError}}
  \]

- When the null hypothesis is true, the expected value of this ratio is around one
- p-value is the probability of getting an F statistic at least as large or larger…
- When the validity conditions are met, this statistic follows an \( F \) distribution with \( \text{df}_{\text{numerator}} = I - 1 \) and \( \text{df}_{\text{denominator}} = N - I \)

Example: Employment Discrimination cont.
(a) Use JMP or R to carry out an ANOVA on the DisabilityEmployment data.
  JMP: Analyze > Fit Y by X; Means/Anova
  R: `summary(aov(score~disability))`

Verify the MS values we found, the F-statistic, and the p-value. In particular, record the MSE value and the degrees of freedom for error. Explain the “significance codes.”

\[
\text{MSE} = 2.666 = s_p^2 \\
5_p = 1.633 \quad \text{df}=65
\]

When we reject the null hypothesis, we should perform a follow-up (“post-hoc”) analysis to help determine which group means differ. There are several recommendations of procedures (“Statisticians disagree sharply on the right way to handle several simultaneous multiple comparisons”), we will explore just two.

1. Fisher’s Least Significant Difference (LSD)

Uses the familiar form: \( \bar{x}_i - \bar{x}_j \pm \text{critical value} \times \text{standard error, with critical value} t_{n-I} \)

and standard error equal to \( s_p \), the pooled estimate across all the groups in the study.
[Note: This is what the applet does after you check the ANOVA table box.]

\[
\bar{x}_i - \bar{x}_j \pm t_{n-I}s_p \sqrt{1/n_i + 1/n_j}
\]

none - amputee: (-0.7611, 1.7040)
none - crutches: (-2.25, 0.21)
none - hearing: (-0.3826, 2.0826)
none - wheelchair: (-1.68, 0.79)
amputee - crutches: (-2.73, -0.26)*
amputee - hearing: (-0.8540, 1.6111)
amputee - wheelchair: (-2.15, 0.32)
crutches - hearing: (0.6389, 3.1040)*
crutches - wheelchair: (-0.6540, 1.8111)
hearing - wheelchair: (-2.53, -0.06)*

\[
(4.9000 - 4.4286) \pm 1.997(1.63293) \sqrt{1/14 + 1/14} \ (\text{df}=65) \]

vs.

\[
(4.9000 - 4.4286) \pm 2.056(1.669) \sqrt{1/14 + 1/14} \ (\text{df}=26)
\]

\[
(4.05 - 5.34) \pm 1.997(1.63293) \sqrt{1/14 + 1/14} \ (\text{df}=65)
\]
Reasoning behind Fisher’s LSD Method: Because we have already declared there is at least one significant difference, we could argue that we are less concerned with the type I error rate (which assumes the null hypothesis is true), and therefore we don’t make any “adjustment” for the simultaneous intervals. So all that changes from what we would have done in the two-sample case is we use the pooled estimate of the variance in each case (and larger df). One consequence is all the intervals are the same length.

Alternatively, we can take the number of simultaneous confidence intervals into account and focus on controlling the overall confidence level that all intervals contain parameters:

2. Tukey’s Honest Significant Difference (HSD)

Uses the familiar form: \( \bar{x}_i - \bar{x}_j \pm \text{critical value} \times \text{standard error} \), where the critical value comes from a different distribution (“studentized range distribution”) which depends on the number of groups you are comparing as well as the error df. This method is a bit more “conservative” but you set the family-wise error rate: “I’m 95% confident that all these intervals simultaneously capture the respective difference in population means…”

(b) Use JMP or R to calculate the Tukey HSD intervals.

JMP: Compare Means > All Pairs, Tukey HSD
R: > PostHocTest(model, method="hsd")

Which pairs of means do you find to be significantly different? How are you/technology deciding? Interpret the results.

(c) How do these intervals compare to the LSD intervals? Consequences?

HSD intervals tend to be wider.

“I’m 98% confident that all these intervals simultaneously capture their parameter.”

One way to visually summarize the results is to sort the means and then draw lines over the pairs that are not significantly different (you can then delete lines that are subsets of other lines). Then summarize which means are not “connected.”

4.05 4.43 4.90 5.34 5.92
hear ampu none wheel crut

Note: There will often be some “ambiguity” in the results! Try to summarize your conclusions as succinctly as possible.
Winter, 2017  Thursday, Feb. 2

```
$disability
diff lwr.ci upr.ci pval
hearing-amputee   -0.3785714 -1.6111894  0.8540465 0.5418
none-hearing        0.8500000 -0.3826179  2.0826179 0.1732
wheelchair-hearing  1.2928571  0.0602392  2.5254751 0.0401 *
hearing-crutches    -1.8714286 -3.1040465 -0.6388106 0.0035 **
none-amputee         0.4714286 -0.7611894  1.7040465 0.4477
wheelchair-amputee  0.9142857 -0.3826179  2.1869037 0.1433
crutches-amputee    1.4928571  0.0602392  2.5254751 0.0184 *

none-crutches       -1.0214286 -2.2540465  0.2111894 0.1028
wheelchair-none    0.4428571 -0.7897608  1.6754751 0.4756
wheelchair-crutches 0.5785714 -1.1111894  0.6540465 0.3520
...

Fisher Pairwise Comparisons

Grouping Information Using the Fisher LSD Method and 95% Confidence

<table>
<thead>
<tr>
<th>disability</th>
<th>N</th>
<th>Mean</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>crutches</td>
<td>14</td>
<td>5.921</td>
<td>A</td>
</tr>
<tr>
<td>wheelchair</td>
<td>14</td>
<td>5.343</td>
<td>A B</td>
</tr>
<tr>
<td>none</td>
<td>14</td>
<td>4.900</td>
<td>A B C</td>
</tr>
<tr>
<td>amputee</td>
<td>14</td>
<td>4.429</td>
<td>B C</td>
</tr>
<tr>
<td>hearing</td>
<td>14</td>
<td>4.050</td>
<td></td>
</tr>
</tbody>
</table>

Means that do not share a letter are significantly different.

Grouping Information Using the Tukey Method and 95% Confidence

<table>
<thead>
<tr>
<th>disability</th>
<th>N</th>
<th>Mean</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>crutches</td>
<td>14</td>
<td>5.921</td>
<td>A</td>
</tr>
<tr>
<td>wheelchair</td>
<td>14</td>
<td>5.343</td>
<td>A B</td>
</tr>
<tr>
<td>none</td>
<td>14</td>
<td>4.900</td>
<td>A B</td>
</tr>
<tr>
<td>amputee</td>
<td>14</td>
<td>4.429</td>
<td>A B</td>
</tr>
<tr>
<td>hearing</td>
<td>14</td>
<td>4.050</td>
<td>B</td>
</tr>
</tbody>
</table>

Means that do not share a letter are significantly different.