Bonferroni Simultaneous C.I.

As with the Tukey adjustment, the Bonferroni method can be used to construct simultaneous C.I.'s for pairwise comparisons (a probability that 1 or more of the true mean differences is not covered by C.I.).

Bonferroni:

\[ \bar{Y}_j - \bar{Y}_k \pm t_{(2m, df)} \frac{MSE(\frac{1}{n_j} + \frac{1}{n_k})}{\sqrt{m(1 + m/d)}} \]

Adjustment based on # of comparisons m will make C.I. wider than per comparison C.I. “df” coincides with df for error.

Tukey C.I.'s will be more narrow than Bonferroni C.I.

Graphical display of pairwise comparisons results

Consider a 5 treatment 1-way ANOVA with equal sample sizes, and these results from a Tukey adjustment:

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.0008</td>
<td>.1111</td>
<td>.0415</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.0415</td>
<td>.1111</td>
<td>.0112</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.9648</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>.0003</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which means are statistically significantly different?
1 is diff than all but 3.
2 is diff than all but 4.
3 not diff than 1 or 4, but is diff than 2 & 5.
4 is not diff than 2 or 3, but is diff than 1 & 5.
5 is different than all others.

Maybe a graphic is useful...

\[ \bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \bar{Y}_4, \bar{Y}_5. \]

10 15 20 25 30

[Lines] join means that are not significantly different.

SAS will do this for you... but be careful using this graphic if you have unequal sample sizes.

In the case of unequal sample sizes, means that are closer together could be significant while those that are farther apart could be not significant (confusing to the client and it doesn't go with the graphic.)
One-way ANOVA with 5 treatment groups

Showing 'means' and 'LSMeans' statement for lines diagram from pairwise comparisons tests.

Three EUs are randomly assigned to each treatment. This is a completely randomized design (CRD).

data mine;
input group y;
cards;
  1 12
  1 17
  1 13
  2 22
  2 24
  2 25
  3 18
  3 17
  3 20
  4 17
  4 21
  4 20
  5 30
  5 29
  5 32;

SAS code and output using 'means' statement:

/*Fit the 1-way ANOVA model and use the 'means' statement to get the
  tukey adjustment, plus get the lines diagram.*/

proc glm data=mine plot=diagnostics;
  class group;
  model y=group;
  means group/tukey lines;     /* Don't put "Adjust=..." in the 'means' statement.*/
run;

The GLM Procedure
Dependent Variable:  y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>459.06666667</td>
<td>114.76666667</td>
<td>31.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>36.66666667</td>
<td>3.66666667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>495.73333333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The GLM Procedure
Tukey’s Studentized Range (HSD) Test for y

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

\[
\begin{align*}
\text{Alpha} & \quad 0.05 \\
\text{Error Degrees of Freedom} & \quad 10 \\
\text{Error Mean Square} & \quad 3.666667 \\
\text{Critical Value of Studentized Range} & \quad 4.65429 \\
\text{Minimum Significant Difference} & \quad 5.1455
\end{align*}
\]

\[
HSD = q_{0.05}(5,10) \cdot \sqrt{\frac{MSE}{n}}
\]

Means with the same letter are not significantly different.

Tukey Grouping

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Mean</th>
<th>N</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30.333</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>23.667</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>19.333</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>18.333</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>14.000</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Letters used to represent lines

ordered means

SAS code and output using `lsmeans` statement:

/* Can get 'lines' diagram in LSmeans, but not Tukey HSD detailed information. */

proc glm data=mine;
  class group;
  model y=group;
  lsmeans group/adjust=tukey pdiff lines; /* Use "Adjust="..." in the 'lsmeans' statement. */
run;

Least Squares Means for effect group
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: y

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.0008</td>
<td>0.1111</td>
<td>0.0415</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.0008</td>
<td></td>
<td>0.0415</td>
<td>0.1111</td>
<td>0.0112</td>
</tr>
<tr>
<td>3</td>
<td>0.1111</td>
<td>0.0415</td>
<td></td>
<td>0.9648</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>0.0415</td>
<td>0.1111</td>
<td>0.9648</td>
<td></td>
<td>0.0003</td>
</tr>
<tr>
<td>5</td>
<td>&lt;.0001</td>
<td>0.0112</td>
<td>0.0001</td>
<td>0.0003</td>
<td></td>
</tr>
</tbody>
</table>
Tukey Comparison Lines for Least Squares Means of group

LS-means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th></th>
<th>LSMEAN</th>
<th>group</th>
<th>LSMEAN Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30.33333</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>23.66667</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>23.66667</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>19.33333</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>18.33333</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>14.00000</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

You will also get a ‘Diffogram’ or ‘mean-mean scatter plot’:

The upward sloping 45-degree dashed line in the plotting area represents points of equality for two means; that is, it shows how close the difference of the two lsmeans is to 0. For this reason both axes have the same length and scale and that all p means are not plotted on each axis, as the difference of any mean with itself is trivially 0 (you will see this point of intersection on the graph for the values of means that appear on both axes). A third axis (implied on the graph but not printed by ODS Graphics) runs at -45 degrees (from the upper left corner of the plotting area to the lower right corner). This axis represents the magnitude of the differences of the lsmeans depicted on the other two axes scaled in such a way that the confidence interval crosses the line of equality when the interval contains 0. It serves as the reference point for the differences between pairs of lsmeans and also gives the approximate values of the endpoints of the confidence interval. Because these data are balanced, the estimated standard errors of all pairwise comparisons are identical, and the widths of the line segments are the same. The labels of the grid lines indicate the ordering of the least squares means.
If the design is unbalanced, you need to be careful using the ‘lines’ graphic, but SAS will automatically correctly apply the Tukey-Kramer adjustment.

Output from the ‘means’ statement when unbalanced:

NOTE: Cell sizes are not equal.

From the SAS documentation on ‘means’:

*Performs Tukey's studentized range test (HSD) on all main-effect means in the MEANS statement. (When the group sizes are different, this is the Tukey-Kramer test.)*

Output from the ‘LSMeans’ statement when unbalanced:

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

From the SAS documentation on ‘LSMeans’:

*When you specify ADJUST=TUKEY and your data are unbalanced, PROC GLM uses the approximation described in Kramer (1956) and identifies the adjustment as “Tukey-Kramer” in the results.*
REGWR (or REGWQ) adjustment for pairwise comparisons

Tukey is probably the most well-known and most used "all-pairwise comparison" adjustment, but there is another that may give more significant comparisons while controlling the strong FWER.

Ryan-Einot-Gabriel-Welsch Range test (REGWR)

This is a "step-down procedure", where we start with the most extreme difference and work our way through smaller differences (if needed).

We will use the data from the previous example to show how the REGWR adjustment works.

<table>
<thead>
<tr>
<th>Observed means</th>
<th>&quot;Ordered&quot; means</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{Y}_{1.}$ = 14.00</td>
<td>$\bar{Y}_{(1)}.$ = 14.00</td>
</tr>
<tr>
<td>$\bar{Y}_{2.}$ = 23.67</td>
<td>$\bar{Y}_{(2)}.$ = 18.33</td>
</tr>
<tr>
<td>$\bar{Y}_{3.}$ = 18.33</td>
<td>$\bar{Y}_{(3)}.$ = 19.33</td>
</tr>
<tr>
<td>$\bar{Y}_{4.}$ = 19.33</td>
<td>$\bar{Y}_{(4)}.$ = 23.67</td>
</tr>
<tr>
<td>$\bar{Y}_{5.}$ = 30.33</td>
<td>$\bar{Y}_{(5)}.$ = 30.33</td>
</tr>
</tbody>
</table>

REGWR procedure focuses on ordered means.
Begin with the most extreme pair (1) and (5).

Test $H_0$ s.t. all the means from (1) up to (5) are equal.
If we fail to reject, we're done and we stop.

If we reject, we declare (1) & (5), the extremes, to be different and we move to the next step...

We now consider 2 "inner stretches" of means from (1) to (4) and from (2) to (5).

$$\overline{\bar{Y}_{(1) to (4)}} \quad \overline{\bar{Y}_{(2) to (5)}}$$

If $H_0$ s.t. all means from (1) to (4) are equal and $H_0$ s.t. all means from (2) to (5) are equal both fail to reject, then we're done and we stop.

Otherwise, we continue the procedure for smaller and smaller "stretches" of means until we no longer reject.
The critical value for testing a stretch of length \( k \) depends on \([n_{i}], \text{ df for error}, \) and \( a \) the # of groups.

\[
U = q_{k,a} \left( \frac{k, df}{\sqrt{2}} \right) \sqrt{2} \quad \text{for} \quad k = a, a-1
\]

"for first & second steps"

\[
U = q_{k,a} \left( \frac{k, df}{\sqrt{2}} \right) \sqrt{2} \quad \text{for} \quad k = a-2, a-3, \ldots, 2.
\]

Increasingly smaller as we move to smaller stretches.

For a difference to be significant between means, we need

\[
| \bar{Y}_{j} - \bar{Y}_{k} | > U \cdot \text{MSE} \cdot \sqrt{\frac{1}{n_{j}} + \frac{1}{n_{k}}}
\]

Our example: Compare means (1) to (5) first with \( k = 5 \)

\[
| \bar{Y}_{(1)} - \bar{Y}_{(5)} | = 16.33
\]

\[
U \cdot \text{MSE} \cdot \sqrt{\frac{1}{n_{(1)}} + \frac{1}{n_{(5)}}} = q_{a} \left( 5, 10 \right) \sqrt{36.666 \gamma \frac{1}{3} + \frac{1}{3}}
\]

\[
= 5.15 < 16.33
\]

\[
\Rightarrow \text{Declare (1) & (5) significantly different.}
\]

\[
\Rightarrow \text{Move to step 2.}
\]
Step #2  stretches of means of length $k=4$.

\[ |\overline{Y}_{(3)} - \overline{Y}_{(4)}| = 9.67 \quad |\overline{Y}_{(2)} - \overline{Y}_{(5)}| = 12.00 \]

\[ t_{0.05}(4,10), \sqrt{\text{MSE} \gamma \frac{1}{{\frac{1}{3}} + \frac{1}{3}}} \approx 4.79 < 9.67 \quad \text{&} \quad 4.79 < 12.00 \]

\[ \Rightarrow \text{Declare (1) & (4) and (2) & (5) sign. diff.} \]

\[ \Rightarrow \text{Move to Step 3} \]

Step #3  stretches of means of length $k=3$.

Since both of the $k=4$ stretches were significant, we have 3 stretches of length $k=3$ to consider.

\[ |\overline{Y}_{(1)} - \overline{Y}_{(3)}| = 5.33 \]

\[ |\overline{Y}_{(2)} - \overline{Y}_{(4)}| = 5.34 \]

\[ |\overline{Y}_{(3)} - \overline{Y}_{(5)}| = 11.00 \]

\[ t_{0.05}(3,10), \sqrt{\text{MSE} \gamma \frac{1}{{\frac{1}{3}} + \frac{1}{3}}} \approx 4.78 \]

\[ \Rightarrow \text{all are significant} \]

\[ \Rightarrow \text{move to Step 4.} \]
Step 4: Stretches of means of length k = 2

We still have all stretches of length 2 to consider

\[ |\bar{Y}_{(1)} - \bar{Y}_{(2)}| = 4.33 \]
\[ |\bar{Y}_{(3)} - \bar{Y}_{(4)}| = 4.34 \]
\[ |\bar{Y}_{(2)} - \bar{Y}_{(3)}| = 1.0 \]
\[ |\bar{Y}_{(4)} - \bar{Y}_{(5)}| = 6.66 \]

\[ \frac{q_{.1} (2, 10) \sqrt{\text{MSE}}}{\sqrt{\frac{1}{3} + \frac{1}{3}}} \approx 4.32 \]

\[ \sqrt{\frac{5}{12}} \]

\[ \Rightarrow \text{All significantly different except (2) and (3).} \]

\[ \Rightarrow \text{Done.} \]

End Results:

\[ \begin{array}{c}
\bar{Y}_1 \quad \bar{Y}_3 \quad \bar{Y}_4 \quad \bar{Y}_2 \quad \bar{Y}_5 \\
10 \quad 15 \quad 20 \quad 25 \quad 30
\end{array} \]

Thus, the REGWR controlled the strong FWER, but gave more significant differences (smaller Type II error).

If there are unequal sample sizes, the error rate is approximate.
SAS statements and partial output for 'means' statement with REGWQ adjustment:

```sas
proc glm data=mine;
   class group;
   model y=group;
   means group/regwq;
run;
```

The GLM Procedure

Ryan-Einot-Gabriel-Welsch Multiple Range Test for y

NOTE: This test controls the Type I experimentwise error rate.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Degrees of Freedom</td>
<td>10</td>
</tr>
<tr>
<td>Error Mean Square</td>
<td>3.666667</td>
</tr>
</tbody>
</table>

Number of Means | 2 | 3 | 4 | 5
Critical Range  | 4.3070844 | 4.7628896 | 4.7832114 | 5.1455143

The GLM Procedure

Ryan-Einot-Gabriel-Welsch Multiple Range Test for y

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>N</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30.333</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>23.667</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>19.333</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>18.333</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>14.000</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Only groups 3 & 4 are not statistically different. But REGWR is not for simultaneous confidence intervals.
More multiple testing adjustments (Still OLRT ch. 5)

the LSD is not advised, but it is mentioned in the book. This is the Least Significant Difference method and it basically makes no adjustment.

\[
LSD = t_{q/2,\nu} \sqrt{MSE} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}
\]

If \(|\bar{y}_i - \bar{y}_j| > LSD\), you can consider \(y_i\) & \(y_j\) significantly different. They call this a "Protected LSD" if you only consider comparisons after rejecting the overall F-test, but no Family-wise error rate is controlled.

What about other follow-up tests besides "all pairwise comparisons"?

⇒ The Scheffe Method for all contrasts...

This will provide an α-level Family-wise error rate for any number of contrasts, but it has rather low power.
For each contrast \( T_u = \sum_{i=1}^{a} C_{ui} M_i \),

\[ = C_{u1} M_1 + C_{u2} M_2 + \ldots + C_{ua} M_a \]

Compute \( \hat{T}_u = \sum_{i=1}^{a} C_{ui} \bar{Y}_i \) as the estimate for \( T_u \),

and the standard error for \( \hat{T}_u \) as

\[ \sqrt{\text{Var}(\hat{T}_u)} = \sqrt{\text{MSE} \sum_{i=1}^{a} \frac{C_{ui}^2 \bar{Y}_i^2}{n_i}}. \]

If \( |\hat{T}_u| > -\sqrt{\text{MSE} \sum_{i=1}^{a} \frac{C_{ui}^2 \bar{Y}_i^2}{n_i}} \cdot \sqrt{(a-1) F_{q, q-1, N-a}} \)

then reject \( H_0: T_u = 0 \). This will control the Family-wise error rate at the \( \alpha \)-level.

You can also use Scheffe method to create simultaneous confidence intervals.

I think of the Scheffe adjustment when you're "snooping around" the data using contrasts.
Comparison with Control (the Dunnett procedure)

there are \((a-1)\) tests of interest.

Simultaneous \((-q)\) confidence intervals are

\[
\overline{Y}_i - \overline{Y}_{\text{CONTROL}} = d.f_{\text{error}} \frac{q(a-1, df)}{MSE} \sqrt{\frac{1}{n_i} + \frac{1}{n_{\text{CONTROL}}}}
\]

and \(d.f_{\text{error}}\) for all \(i \neq \text{control}\)

and \(q(a-1, df)\) can be found in the D.9 tables in the back of OLRT pp. 635-638.

Or, we can consider Dunnett's significant difference as

\[
\text{DSD} = q(a-1, df) \sqrt{\frac{1}{n_i} + \frac{1}{n_{\text{CONTROL}}}}
\]

the procedure controls the strong FWER.

(it is approximate if not equal sample sizes.)

Because the control group is used in every comparison, it is a good idea to give greater replication to the control group.

It turns out for a fixed total sample size \(N\), the way to minimize the average variance of all \(\overline{Y}_i - \overline{Y}_{\text{CONTROL}}\) would be to set the ratio \(\frac{n_{\text{CONTROL}}}{n_i} = 1\).
Quick Guideline:

**Situation**

All pairwise comparisons

**Procedure**

- balanced data (Tukey or REGWQ)
- unbalanced data (Tukey-Kramer)
- Scheffé

All possible contrasts

\[ \lambda_1 = \frac{1}{2} (\lambda_2 + \lambda_3) \]

Many linear combinations

\[ \lambda_1 + \lambda_2 + \lambda_3 = 0 \]

Comparisons with control

Bonferroni

(Always be applied)

Dunnett

*All of these control the strong FWER.*