Tukey’s multiple comparisons adjustment - revisited

STAT:5201

Week 13 - Lecture 1
Tukey’s adjustment in a 1-way ANOVA

1-way ANOVA (1 factor factorial, CRD)

- $n$ observations in each cell (balanced)
- $g$ groups (or treatments)
- Use Tukey’s adjustment to control the FWER at the $\alpha$ level for all pairwise comparisons

$$HSD = \frac{q_{\alpha(g,\nu)}}{\sqrt{2}} S.E. (\bar{Y}_i - \bar{Y}_j)$$

$$= \frac{q_{\alpha(g,\nu)}}{\sqrt{2}} \sqrt{MSE(\frac{1}{n} + \frac{1}{n})}$$

$$= q_{\alpha(g,\nu)} \sqrt{MSE(\frac{1}{n})}$$

- where $\nu$ is degrees of freedom for $MSE$, in this case $\nu = gn - g$
- $\mu_i$ and $\mu_j$ declared significantly different if $|\bar{Y}_i - \bar{Y}_j| > HSD$
Tukey’s adjustment in a 2-way ANOVA, no interaction

2-way ANOVA (2 factor factorial, CRD)

Here, we will assume we have a main-effects only model

- $a$ levels of factor A
- $b$ levels of factor B
- $n$ observations in each cell (balanced)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \text{with} \quad \epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$
Tukey’s adjustment in a 2-way ANOVA, no interaction

Tukey’s adjustment for the ‘main effects’:

- Tukey HSD for comparing factor A means (there are $a$ means)

$$HSD_A = \frac{q_{\alpha(a,\nu)}}{\sqrt{2}} S.E.(\bar{Y}_{i..} - \bar{Y}_{j..})$$

$$= \frac{q_{\alpha(a,\nu)}}{\sqrt{2}} \sqrt{MSE(\frac{1}{bn} + \frac{1}{bn})}$$

where $\nu$ is degrees of freedom for $MSE$, in this case $\nu = abn - a - b + 1$

- Use Tukey’s adjustment to control the FWER of all pairwise comparisons within factor A at the $\alpha$ level

- $\mu_i.$ and $\mu_j.$ declared significantly different if $|\bar{Y}_{i..} - \bar{Y}_{j..}| > HSD_A$
Tukey’s adjustment in a 2-way ANOVA, no interaction

Tukey’s adjustment for the ‘main effects’:

- Tukey HSD for comparing factor B means (there are $b$ means)

$$HSD_B = \frac{q_{\alpha(b, \nu)}}{\sqrt{2}} S.E.(\bar{Y}_{i..} - \bar{Y}_{j..})$$

$$= \frac{q_{\alpha(b, \nu)}}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{an} + \frac{1}{an}\right)}$$

- where $\nu$ is degrees of freedom for $MSE$, in this case $\nu = abn - a - b + 1$

- Use Tukey’s adjustment to control the FWER of all pairwise comparisons within factor B at the $\alpha$ level

- $\mu_i$ and $\mu_j$ declared significantly different if $|\bar{Y}_{i..} - \bar{Y}_{j..}| > HSD_B$

NOTE: If you have significant AB interaction, you can think of the 2 factors combining into a single superfactor and approach it as a 1-way ANOVA with $ab$ groups.
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

Hyvee is interested in determining the best cake mix to use for their locally baked strawberry shortcake.

- There are 3 brands currently available on the shelf and they decide the easiest method would be to evaluate each of these to determine which is best.
- To have a good representative sample they decide to make 4 boxes from each brand.
- Additionally, they decide that they want to try two temperatures (low/high) for each brand.
- To get the most info from their boxes, they will split each box into 2 separate cakes, one for each temperature, for a total of 24 cakes [3 brands × 4 boxes × 2 temperatures].
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

The Lenth diagram:

![Lenth Diagram]

\[ \text{brand}^3 \times \text{temp}^2 \]

\[ \text{BOX}^{12} \]

```plaintext
proc mixed data=data plots(only)=Residualpanel(conditional);
  class Brand Box Temp;
  model Moisture=Temp Brand Temp*Brand/ddfm=satterth;
  random Box(Brand);
run;
```
Tukey’s adjustment in a Type I split plot design, no interaction

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>1</td>
<td>9</td>
<td>9.68</td>
<td>0.0125</td>
</tr>
<tr>
<td>Brand</td>
<td>2</td>
<td>9</td>
<td>6.68</td>
<td>0.0166</td>
</tr>
<tr>
<td>Brand*Temp</td>
<td>2</td>
<td>9</td>
<td>1.36</td>
<td>0.3054</td>
</tr>
</tbody>
</table>

- There is no significant interaction.
- There is a significant $Temp$ effect.
- There is a significant $Brand$ effect and we wish to compare the levels of Brand (i.e. factor A) and control the FWER at the $\alpha$ level.
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

- There are 3 brands leading to \( \binom{3}{2} = 3 \) brand comparisons.
- The within-box factor is temp.
- The between-box factor is brand.
- Let \( Y_{ijk} \) represent the response for box \( j \) nested in brand \( i \) at temp \( k \), with \( i = 1, 2, 3 \) and \( j = 1, 2, 3, 4 \) and \( k = 1, 2 \).
- The S.E. for comparing brand 1 and brand 2:

\[
\text{Var}(\bar{Y}_{i..} - \bar{Y}_{j..}) = \frac{1}{4}(\sigma^2 + 2\sigma_{\text{box(brand)}}^2)
\]

Estimated: \( S.E.(\bar{Y}_{i..} - \bar{Y}_{j..}) = \sqrt{\frac{1}{4} MS_{\text{box(brand)}}} \)
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

- Tukey’s HSD for comparing the \(a = 3\) brands (main effects):

\[
HSD_{Brand} = \frac{q_{\alpha(3,\nu)}}{\sqrt{2}} \cdot S.E.(\bar{Y}_{i.} - \bar{Y}_{j.})
\]

\[= \frac{q_{\alpha(3,\nu)}}{\sqrt{2}} \sqrt{\frac{1}{4} MS_{box(brand)}}
\]

- where \(\nu\) is degrees of freedom coinciding with \(MS_{box(brand)}\), in this case \(\nu = 9\).

- Two brands are declared significantly different if

\[|\bar{Y}_{i.} - \bar{Y}_{j.}| > HSD_{Brand}\]
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

The GLM Procedure
Dependent Variable: Moisture

<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>14</td>
<td>1513.870623</td>
<td>108.133616</td>
<td>2.91</td>
<td>0.0561</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>334.775144</td>
<td>37.197238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23</td>
<td>1848.645766</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>1</td>
<td>360.1947767</td>
<td>360.1947767</td>
<td>9.68</td>
<td>0.0125</td>
</tr>
<tr>
<td>Box(Brand)</td>
<td>9</td>
<td>423.6632448</td>
<td>47.0736939</td>
<td>1.27</td>
<td>0.3657</td>
</tr>
<tr>
<td>Brand</td>
<td>2</td>
<td>629.0528931</td>
<td>314.5264465</td>
<td>8.46</td>
<td>0.0086</td>
</tr>
<tr>
<td>Brand*Temp</td>
<td>2</td>
<td>100.9597082</td>
<td>50.4798541</td>
<td>1.36</td>
<td>0.3054</td>
</tr>
</tbody>
</table>
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

- \( S.E.(\bar{Y}_{i..} - \bar{Y}_{j..}) = \sqrt{\frac{1}{4} MS_{box(brand)}} \)

\[ = \sqrt{\frac{1}{4} \cdot 47.07} = 3.43 \]

Differences of Least Squares Means (from SAS)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Brand 1</th>
<th>Brand 2</th>
<th>Brand 3</th>
<th>Estimate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand 1</td>
<td>1</td>
<td>2</td>
<td></td>
<td>9.8538</td>
<td>3.4305</td>
</tr>
<tr>
<td>Brand 1</td>
<td>1</td>
<td>3</td>
<td></td>
<td>11.6443</td>
<td>3.4305</td>
</tr>
<tr>
<td>Brand 2</td>
<td>2</td>
<td>3</td>
<td></td>
<td>1.7905</td>
<td>3.4305</td>
</tr>
</tbody>
</table>
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

proc mixed data=data plots(only)=Residualpanel(conditional);
    class Brand Box Temp;
    model Moisture=Temp Brand Temp*Brand/ddfm=satterth;
    random Box(Brand);
    lsmeans Brand/adjust=Tukey pdiff;
run;

Least Squares Means

| Effect | Brand | Estimate | Standard Error | DF | t Value | Pr > |t| |
|--------|-------|----------|----------------|----|---------|------|---|
| Brand  | 1     | 61.5457  | 2.4257         | 9  | 25.37   | <.0001|
| Brand  | 2     | 51.6918  | 2.4257         | 9  | 21.31   | <.0001|
| Brand  | 3     | 49.9013  | 2.4257         | 9  | 20.57   | <.0001|
Tukey’s adjustment in a Type I split plot design, no interaction

Example (Best cake mix)

\[ HSD_{Brand} = \frac{q_{\alpha(3,9)}}{\sqrt{2}} \cdot 3.43 = \frac{3.95}{\sqrt{2}} \cdot 3.43 = 9.58 \]

Differences of Least Squares Means

| Effect  | Brand | Brand | Estimate | Error  | DF | t Value | Pr > |t| | Adjustment |
|---------|-------|-------|----------|--------|----|---------|------|---|------------|
| Brand 1 | 1     | 2     | 9.8538   | 3.4305 | 9  | 2.87    | 0.0184|   | Tukey      |
| Brand 1 | 1     | 3     | 11.6443  | 3.4305 | 9  | 3.39    | 0.0079|   | Tukey      |
| Brand 2 | 2     | 3     | 1.7905   | 3.4305 | 9  | 0.52    | 0.6143|   | Tukey      |

Brand 1 is significantly different than 2 & 3, but 2 & 3 are not different from each other.
Tukey’s adjustment in a Type I split plot design, no interaction

Re-fit model without interaction (i.e. change mean structure, and put the SS for interaction into the SS for error).

Example (Best cake mix)

```
proc mixed data=data plots(only)=Residualpanel(conditional);
  class Brand Box Temp;
  model Moisture=Temp Brand/ddfm=satterth outpredm=forplotting;
  random Box(Brand);
  lsmeans Brand/adjust=Tukey pdiff;
run;
```

| Effect  | Brand | Temp | Estimate | Error | DF  | t Value | Pr > |t| |
|---------|-------|------|----------|-------|------|---------|-------|---|
| Brand   | 1     |      | 61.5457  | 2.4257| 9    | 25.37   | <.0001|
| Brand   | 2     |      | 51.6918  | 2.4257| 9    | 21.31   | <.0001|
| Brand   | 3     |      | 49.9013  | 2.4257| 9    | 20.57   | <.0001|
| Temp    | 1     |      | 50.5056  | 1.9005| 19.3 | 26.57   | <.0001|
| Temp    | 2     |      | 58.2536  | 1.9005| 19.3 | 30.65   | <.0001|

NOTE: I outputted the marginal means (averaged over random effects) into ‘forplotting’ so I could plot my mean structure in R.
Tukey’s adjustment in a Type I split plot design, no interaction

**Example (Best cake mix)**

- **NOTE:** If the split-plot factor (here, Temp) had multiple levels and you wanted to apply the Tukey adjustment to the main effects for this factor, the S.E. would only relate to $MSE$.

- $HSD_{Temp} = \frac{q_{\alpha}(\# \text{ levels of Temp}, \nu)}{\sqrt{2}} S.E.(\bar{Y}_{..i} - \bar{Y}_{..j}) = \sqrt{MSE(\frac{1}{6})}$
  
  where $\nu$ is the degrees of freedom for $MSE$. 

\[ \text{mean of Pred} \]

\[
\begin{array}{ccc}
\text{Temp} \\
2 & 65 \\
3 & 60 \\
1 & 50 \\
\end{array}
\]

\[
\begin{array}{ccc}
\text{Brand} \\
1 & 55 \\
2 & 60 \\
3 & 50 \\
\end{array}
\]