'Planned' comparisons vs unplanned comparisons

Confidence Intervals Multiple Comparisons: HSD

Remark Alternate form of Model I

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \sum_{i} n_i \alpha_i = 0 \text{ identifiability constraint} \]

'Planned' comparisons - single pairs of means, or 'constraints' specified in advance

Difference of Means e.g. \( \mu_i - \mu_j \) : like a two-sample test
- but, we have an ANOVA model and hence the pooled variance estimate \( s^2 \) for the common variance \( \sigma^2 \).

\[ 100(1-\alpha)\%CI \quad \bar{y}_i - \bar{y}_j \pm t_{1-\alpha/2[\nu]} SE_{\bar{y}_i-\bar{y}_j}, \nu = n - a, SE_{\bar{y}_i-\bar{y}_j} = s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \]

Ex: (Pea section data) Length of pea sections grown in tissue cultures

<table>
<thead>
<tr>
<th></th>
<th>Glucose</th>
<th>Fructose</th>
<th>Gluc+</th>
<th>Fruc</th>
<th>Sucrose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>57</td>
<td>58</td>
<td>58</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>58</td>
<td>61</td>
<td>59</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>60</td>
<td>56</td>
<td>58</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>59</td>
<td>58</td>
<td>61</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>62</td>
<td>57</td>
<td>57</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>60</td>
<td>56</td>
<td>56</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>60</td>
<td>61</td>
<td>58</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>57</td>
<td>60</td>
<td>57</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>59</td>
<td>57</td>
<td>57</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>61</td>
<td>58</td>
<td>59</td>
<td>67</td>
<td></td>
</tr>
</tbody>
</table>

peas=scan()
pea.df=data.frame(peas,culture=as.factor(rep(1:5,10)))
culture=as.factor(rep(1:5,10))
culture

[1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5

Levels: 1 2 3 4 5
pea.lm=lm(peas~culture,data=pea.df)
anova(pea.lm)
Response: peas

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>culture</td>
<td>4 1077.32 269.33 49.368 6.737e-16 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>45 245.50 5.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
pea.resid=residuals(pea.lm)
pea.fitted=fitted(pea.lm)
matrix(round(pea.resid,2),ncol=5,byrow=T)

\[ \begin{bmatrix}
[1,] & 4.9 & -2.3 & -0.2 & 0 & -2.1 \\
\end{bmatrix} \]
95% CI for $\mu_c - \mu_g$, difference between the control group and the glucose group: $\bar{y}_c - \bar{y}_g = 70.1 - 59.3 = 10.8$

$$s = \sqrt{s^2} = \sqrt{MS_{\text{within}}} = \sqrt{5.46} = 2.34$$

$$SE_{\bar{y}_c - \bar{y}_g} = 2.34 \sqrt{\frac{1}{10} + \frac{1}{10}} = 1.046$$

dof $\nu = n - a = 50 - 5 = 45$; $t_{.975[45]} = 2.015$

C.I. = 10.8 ± (2.02)(1.046) = [8.69, 12.91]

**Contrast**: A linear combination of means where the coefficients sum to zero: Population

$$\gamma = \sum_{i=1}^{a} c_i \mu_i \sum_{i=1}^{a} c_i = 0$$

Sample

$$\sum_{i} c_i \bar{y}_i$$

Ex: ‘sugars vs. control’ $\gamma = \mu_c - \frac{1}{4}(\mu_g + \mu_f + \mu_{g+f} + \mu_s)$, coefficients $(c_i) = (1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4})$

Typically used to compare groups of means or certain weighted combinations (‘orthogonal contrasts’) such as ‘linear’ or ‘quadratic’ effects.

Variance of a sample contrast (assuming the sample means are independent)

$$\text{Var } c = \text{Var} \left( \sum c_i \bar{y}_i \right) = \sum c_i^2 \text{Var} (\bar{y}_i) = \sum c_i^2 \frac{\sigma^2}{n_i}$$

Estimated $SE_c, SE_c = s \sqrt{\sum_i \frac{c_i^2}{n_i}}$,

$$100(1 - \alpha)\% CI \quad C \pm t_{1-\alpha/2}[\nu]SE_c$$
Ex: \( C = 70.1 - \frac{1}{3}(59.3 + 58.2 + 58 + 64.1) = 10.2 \)

\[
\sum \frac{c_i^2}{n_i} = \frac{1}{10} [1 + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}] = \frac{5}{40} = \frac{1}{8}, \quad \sqrt{\sum \frac{c_i^2}{n_i}} = \sqrt{\frac{1}{8}} = 0.3536
\]

\( SE_c = (2.34) \times (0.3536) = 0.8273 \)

CI for C is 10.2 \( \pm \) (2.02) \times (0.8273) = [8.53, 11.87]

Hypothesis Test for Contrast: e.g. \( H_0 : \gamma = \gamma_0 \)

Form a t-statistic

\[
t = \frac{C - \gamma_0}{SE_c} \sim t_\nu \quad \text{if} \ H_0 \ \text{is true.}
\]

Remark: differences between means are a special case of contrasts: e.g. \( \mu_c - \mu_g = \sum c_i \mu_i \) with \( (C_i) = (1, -1, 0, 0, 0) \).

These types of investigations should be done on combinations of factors that were determined in advance of observing the experimental results, or else the confidence levels are not as specified by the procedure. Also, doing several comparisons might change the overall confidence level. This can be avoided by carefully selecting contrasts to investigate in advance and making sure that:

- the number of such contrasts does not exceed the number of degrees of freedom between the treatments *only or-
- the number of such contrasts does not exceed the number of degrees of freedom between the treatments *only orthogonal contrasts are chosen.

However, there are also several powerful multiple comparison procedures we can use after observing the experimental results.

**Unplanned comparisons** After looking at the data, we may wish to assess the significance of, or give C.I.’s for certain differences e.g. \( \mu_g - \mu_f \) (in pea length e.g.) or contrasts e.g. \( \mu_a - \frac{1}{3} (\mu_g + \mu_f + \mu_gf) \)

That looked interesting a posteriori. Viewed a priori, however, there are many differences or contrasts that could potentially attract attention. We need to adjust our significance levels and P-values (larger) or our C.I.’s (wider) to allow for this search over all possibilities.

**Subject of multiple comparisons**

-see books, e.g. Miller, R.G. Simulatenous Statistic Inference

All pairs of differences with \( a \) treatments, there are \( \binom{a}{2} = \frac{a(a-1)}{2} \) possible comparisons of different means: \( \mu_i - \mu_j, i = 1, \ldots, a; j = 1, \ldots, i - 1 \)

If we used t-intervals, would have many intervals of form

\[
I_{ij} \leftrightarrow \bar{y}_i - \bar{y}_j \pm t_{1-\alpha/2|\nu|} SE_{\bar{y}_i - \bar{y}_j}
\]

But the chance that all intervals simultaneously cover all \( \mu_i - \mu_j \):

\[
P\{I_{ij} \text{ cover} \mu_i - \mu_j; \text{ for all} i < j\} < 1 - \alpha
\]

To obtain a simultaneous coverage property, make intervals wider

\[
I_{ij}^{TK} \leftrightarrow \bar{y}_i - \bar{y}_j \pm Q_{1-\alpha[a,a]} SE_{\bar{y}_i - \bar{y}_j}
\]

TK = 'Tukey Kramer’ \( Q_{1-\alpha[a,a]} \) are percentage points of studentized range distribution.

Formal definition:

\[
Q_{[a,a]} = \max \frac{|Z_i - Z_j|}{s}
\]

where \( Z_1, Z_2, \ldots, Z_a \sim N(0, 1); \nu s^2 \sim \chi^2(\nu) \) and all independent gives wider intervals \( \frac{Q_{1-\alpha[a,a]}}{\sqrt{2}} > t_{1-\alpha/2|\nu|} \) (unless \( a=2! \))

Ex. (pea lengths)

\[
\frac{Q_{.95[5,45]}}{\sqrt{2}} = \frac{4.02}{\sqrt{2}} = 2.84 (> 2.02 = t_{.975[45]})
\]
simultaneous interval for $\mu_c - \mu_g$ is $10.8 \pm (2.843)(1.046) = [7.83, 13.8]$.
Simultaneous coverage property if Model I holds, and $n_1 = n_2 = \ldots = n_a$ ('balanced'), then $P(I_{ij}^{TK}$ covers $\mu_i - \mu_j$ for all $i < j) = 1 - \alpha$

Remark: If the ANOVA is unbalanced (not all $n_i$ equal) then these Tukey-Kramer intervals are conservative (coverage prob $\geq 1 - \alpha$). When comparing the means for the levels of a factor in an analysis of variance, a simple comparison using t-tests will inflate the probability of declaring a significant difference when it is not in fact present. This because the intervals are calculated with a given coverage probability for each interval but the interpretation of the coverage is usually with respect to the entire family of intervals. John Tukey introduced intervals based on the range of the sample means rather than the individual differences. The intervals returned by this function are based on this Studentized range statistics. Technically the intervals constructed in this way would only apply to balanced designs where there are the same number of observations made at each level of the factor. This function incorporates an adjustment for sample size that produces sensible intervals for mildly unbalanced designs.

The term 'experiment wise error rate $\alpha$ ' arises because, if $H_0$ is true (all $\mu_i$ equal), then the chance of falsely declaring as significant any of the $\frac{a(a-1)}{2}$ pair wise diffs is (at most) $\alpha$:

$$P_{H_0}\left\{ \max \frac{|\bar{y}_i - \bar{y}_j|}{SE_{\bar{y}_i - \bar{y}_j}} > \frac{Q_{1-\alpha/2[a,\nu]}}{\sqrt{2}} \right\} \leq \alpha$$

(= $\alpha$ if all $n_i$ equal)

**Balanced Case and Honestly Significant Difference (HSD)**

if all $n_i = n$, then all $SE_{\bar{y}_i - \bar{y}_j} = \frac{s}{\sqrt{n}}$

$\iff \bar{y}_i - \bar{y}_j > Q_{1-\alpha/2[a,\nu]} \frac{s}{\sqrt{n}}$ so just find those pairs ($\bar{y}_i - \bar{y}_j$) separated by $> \text{HSD}$
- the $\pm \frac{1}{2} HSD$ intervals overlap if and only if $|\bar{y}_i - \bar{y}_j| \leq HSD \Rightarrow$ means ($\mu_i; \mu_j$) whose $\frac{1}{2} HSD$ intervals don't overlap are significantly different at experiment wise error rate $\alpha$.

(Warning! $\bar{y}_i \pm \frac{1}{2} HSD$ is NOT a $100(1 - \alpha)\%$ Conf. interval!)

**All contrasts:** The Scheffé intervals

$$I^s \leftrightarrow \pm \sqrt{(a - 1)F_{1-a,\nu}}SE_c$$

have the simultaneous coverage property (for balanced or unbalanced cases)

$$P\{I^s \text{ cover for all contrasts} \} = a - \alpha$$

Since contrasts are more general than differences, expect Scheffé intervals to be even wider than Tukey-Kramer

$\gamma = \mu_s - \frac{1}{3}(\mu_g + \mu_f + \mu_{gf})c = (0, 1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}) \sum \frac{c_i^2}{n_i} = \frac{1}{10}(1 + 3 + \frac{1}{3}) = \frac{4}{32}$

$$SE_c = s\sqrt{\frac{c^2}{n}} = (2.34)(.365) = 0.8544 \sqrt{(a - 1)F_{1-a,\nu}} = \sqrt{4F_{.95}[4,45]} = \sqrt{4x(2.58)} = 3.21 \text{ 95\% Scheffé interval for } c = \bar{x}_s - \frac{1}{3}(\bar{x}_g + \bar{x}_f + \bar{x}_{gf}) = 64.1 - 58.5 = 5.6 \text{ has margin effor } (3.21)(0.8544) = 2.743$

CI [5.6 - 2.74, 5.6 + 2.74] = [2.86, 8.34]

(Note that Scheffé multiplier = 3.21 > 2.84 = $Q_{\alpha[a,\nu]}\sqrt{2} = $ Tukey-Kramer multiplier)

Remark:
- There is a version of the T-K intervals for contrasts
- these can be better (shorter) than the Scheffé method if $a$ is larger and relatively fewer $c_i$ are non-zero

contr.peas=matrix(c(4,-1,-1,-1,-1,0,-1,-1,3,-1),ncol=2)
contr.peas
[,1] [,2]
[1,] 4 0
[2,] -1 -1
[3,] -1 -1
[4,] -1 3
[5,] -1 -1
contrasts(culture)=contr.peas