STAT 222 Lecture 15-16
Contrasts & Multiple Comparisons for Factorial Data

Yibi Huang

Coverage

- Section 6.3 Contrasts
- Section 6.4.3 Multiple Comparisons
- Section 6.5.6 Model Building
- Section 7.4 A Real Experiment—Popcorn–Microwave Experiment

Contrasts

Recall in a one-way model

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

a *contrast* is a linear combination of treatment mean μ_i's

$$C = \sum_{i=1}^{g} c_i \mu_i \quad \text{such that} \quad \sum_{i=1}^{g} c_i = 0$$

Similarly in a factorial model, say a 3-way model,

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ijk\ell},$$

a *contrast* is a linear combination of μ_{ijk}'s

$$C = \sum_{ijk} c_{ijk} \mu_{ijk} \quad \text{such that} \quad \sum_{ijk} c_{ijk} = 0$$

Example: Popcorn Microwave Data (Section 7.4, Review)

We'll demonstrate using the Popcorn Microwave Data in Section 7.4 and Slides L1314.pdf

Brand	Power	Time (k)		
(i)	(j)	1 (4 min)	2 (4.5 min)	3 (5 min)
1	1 (500 W)	73.8, 65.5	70.3, 91.0	72.7, 81.9
1	2 (625 W)	70.8, 75.3	78.7, 88.7	74.1, 72.1
2	1 (500 W)	73.7, 65.8	93.4, 76.3	45.3, 47.6
2	2 (625 W)	79.3, 86.5	92.2, 84.7	66.3, 45.7
3	1 (500 W)	62.5, 65.0	50.1, 81.5	51.4, 67.7
3	2 (625 W)	82.1, 74.5	71.5, 80.0	64.0, 77.0

```
popcorn = read.table(
          "http://www.stat.uchicago.edu/~yibi/s222/popcorn.txt", h=T)
```

Contrasts of Interest (1)

Fitting a 3-way model

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ij}, \quad \text{where } i = \texttt{brand},\ j = \texttt{power},\ k = \texttt{time}$$

of the Popcorn Data, we might be interested in

1. Which combination of `brand`, `power`, and `time` will produce the highest popping rate?

2. Which `brand` performs best
   - overall, averaging over levels of `power` & `time`?
   - for `power` = 625W ($j = 2$) and `time` = 4.5 min ($k = 2$)?
   - for `power` = 625W ($j = 2$), averaging over levels of `time`?

Contrasts of Interest (1)

Fitting a 3-way model

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ij}, \quad \text{where } i = \texttt{brand},\ j = \texttt{power},\ k = \texttt{time}$$

of the Popcorn Data, we might be interested in

1. Which combination of `brand`, `power`, and `time` will produce the highest popping rate?
   - pairwise comparisons $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. Which `brand` performs best
   - overall, averaging over levels of `power` & `time`?
   - for `power` = 625W ($j = 2$) and `time` = 4.5 min ($k = 2$)?
   - for `power` = 625W ($j = 2$), averaging over levels of `time`?

Contrasts of Interest (1)

Fitting a 3-way model

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ij}, \quad \text{where } i = \texttt{brand},\ j = \texttt{power},\ k = \texttt{time}$$

of the Popcorn Data, we might be interested in

1. Which combination of `brand`, `power`, and `time` will produce the highest popping rate?
   - pairwise comparisons $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. Which `brand` performs best
   - overall, averaging over levels of `power` & `time`?
     .... $\bar{\mu}_{i_1\bullet\bullet} - \bar{\mu}_{i_2\bullet\bullet}$
   - for `power` = 625W ($j = 2$) and `time` = 4.5 min ($k = 2$)?
   - for `power` = 625W ($j = 2$), averaging over levels of `time`?

Contrasts of Interest (1)

Fitting a 3-way model

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ij}, \quad \text{where } i = \texttt{brand},\ j = \texttt{power},\ k = \texttt{time}$$

of the Popcorn Data, we might be interested in

1. Which combination of `brand`, `power`, and `time` will produce the highest popping rate?
   - pairwise comparisons $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. Which `brand` performs best
   - overall, averaging over levels of `power` & `time`?
     .... $\bar{\mu}_{i_1 \bullet\bullet} - \bar{\mu}_{i_2 \bullet\bullet}$
   - for `power` = 625W ($j = 2$) and `time` = 4.5 min ($k = 2$)?
     .... $\mu_{i_1 22} - \mu_{i_2 22}$
   - for `power` = 625W ($j = 2$), averaging over levels of `time`?

Contrasts of Interest (1)

Fitting a 3-way model

$$y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ij}, \quad \text{where } i = \texttt{brand},\ j = \texttt{power},\ k = \texttt{time}$$

of the Popcorn Data, we might be interested in

1. Which combination of `brand`, `power`, and `time` will produce the highest popping rate?
   - pairwise comparisons $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. Which `brand` performs best
   - overall, averaging over levels of `power` & `time`?
     .... $\bar{\mu}_{i_1\bullet\bullet} - \bar{\mu}_{i_2\bullet\bullet}$
   - for `power` = 625W ($j = 2$) and `time` = 4.5 min ($k = 2$)?
     .... $\mu_{i_1 22} - \mu_{i_2 22}$
   - for `power` = 625W ($j = 2$), averaging over levels of `time`?
     .... $\bar{\mu}_{i_1 2\bullet} - \bar{\mu}_{i_2 2\bullet}$

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brand`s?

4. Whether the `time` effects change over `power`
   - for `brand` 1?
   - averaging over the 3 `brand`s?

5. Whether the `time` effects change over `brand`
   - for `power` = 625W?
   - averaging over levels of `power`?

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brand`s?
   - pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2
4. Whether the `time` effects change over `power`
   - for `brand` 1?
   - averaging over the 3 `brand`s?
5. Whether the `time` effects change over `brand`
   - for `power` = 625W?
   - averaging over levels of `power`?

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brands`?
   - pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2
4. Whether the `time` effects change over `power`
   - for `brand` 1?

     .... $\mu_{i2k_1} - \mu_{i2k_2} - (\mu_{i1k_1} - \mu_{i1k_2})$ for $i = 1$
   - averaging over the 3 `brands`?
5. Whether the `time` effects change over `brand`
   - for `power` $=$ 625W?
   - averaging over levels of `power`?

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brand`s?
   - pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2
4. Whether the `time` effects change over `power`
   - for `brand` 1?

     .... $\mu_{i2k_1} - \mu_{i2k_2} - (\mu_{i1k_1} - \mu_{i1k_2})$ for $i = 1$
   - averaging over the 3 `brand`s?

     .... $\bar{\mu}_{\bullet 2k_1} - \bar{\mu}_{\bullet 2k_2} - (\bar{\mu}_{\bullet 1k_1} - \bar{\mu}_{\bullet 1k_2})$
5. Whether the `time` effects change over `brand`
   - for `power` = 625W?
   - averaging over levels of `power`?

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brand`s?
   - pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2
4. Whether the `time` effects change over `power`
   - for `brand` 1?
     .... $\mu_{i2k_1} - \mu_{i2k_2} - (\mu_{i1k_1} - \mu_{i1k_2})$ for $i = 1$
   - averaging over the 3 `brand`s?
     .... $\bar{\mu}_{\bullet 2k_1} - \bar{\mu}_{\bullet 2k_2} - (\bar{\mu}_{\bullet 1k_1} - \bar{\mu}_{\bullet 1k_2})$
5. Whether the `time` effects change over `brand`
   - for `power` = 625W?
     .... $\mu_{i_1jk_1} - \mu_{i_1jk_2} - (\mu_{i_2jk_1} - \mu_{i_2jk_2})$ for $j = 2$
   - averaging over levels of `power`?

Contrasts of Interest (2)

3. Which `power:time` combination works the best, averaging over the 3 `brand`s?
   - pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2
4. Whether the `time` effects change over `power`
   - for `brand` 1?

     .... $\mu_{i2k_1} - \mu_{i2k_2} - (\mu_{i1k_1} - \mu_{i1k_2})$ for $i = 1$
   - averaging over the 3 `brand`s?

     .... $\bar{\mu}_{\bullet 2k_1} - \bar{\mu}_{\bullet 2k_2} - (\bar{\mu}_{\bullet 1k_1} - \bar{\mu}_{\bullet 1k_2})$
5. Whether the `time` effects change over `brand`
   - for `power` $=$ 625W?

     .... $\mu_{i_1jk_1} - \mu_{i_1jk_2} - (\mu_{i_2jk_1} - \mu_{i_2jk_2})$ for $j = 2$
   - averaging over levels of `power`?

     .... $\bar{\mu}_{i_1\bullet k_1} - \bar{\mu}_{i_1\bullet k_2} - (\bar{\mu}_{i_2\bullet k_1} - \bar{\mu}_{i_2\bullet k_2})$

Estimation and Standard Error of a Contrast

For example, for 3-way data, a natural estimator for the contrast $C = \sum_{ijk} c_{ijk}\mu_{ijk}$ is

$$\widehat{C} = \sum_{ijk} c_{ijk}\widehat{\mu}_{ijk} = \sum_{ijk} c_{ijk}\overline{y}_{ijk\bullet}$$

By the independence of $\overline{y}_{ijk\bullet}$'s, we know

$$\begin{aligned}\mathbb{V}\Big(\sum_{ijk} c_{ijk}\overline{y}_{ijk\bullet}\Big) &= \sum_{ijk} \mathbb{V}(c_{ijk}\overline{y}_{ijk\bullet}) \\ &= \sum_{ijk} c_{ijk}^2\, \mathbb{V}(\overline{y}_{ijk\bullet}) = \sum_{ijk} c_{ijk}^2 \frac{\sigma^2}{n_{ijk}} .\end{aligned}$$

Here n_{ijk} is the number of replicates in treatment (i, j, k). The estimator $\widehat{C}$ has standard deviation and standard error

$$\mathrm{SD}(\widehat{C}) = \sigma\sqrt{\sum_{ijk} \frac{c_{ijk}^2}{n_{ijk}}}, \qquad \mathrm{SE}(\widehat{C}) = \sqrt{\mathrm{MSE}}\sqrt{\sum_{ijk} \frac{c_{ijk}^2}{n_{ijk}}}$$

Inference for a Single Contrast (No Multiple Comparisons)

A $(1-\alpha)100\%$ confidence interval for a *single* contrast C is

$$\widehat{C} \pm t_{dfE,\alpha/2} \times \mathrm{SE}(\widehat{C})$$

For testing H_0: $C = 0$, the test statistic is

$$t = \frac{\widehat{C}}{\mathrm{SE}(\widehat{C})} \sim t_{dfE}$$

where dfE represents the degrees of freedom for SSE.

Multiple Comparisons for Factorial Data

Just like one-way data, we usually conduct multiple tests and construct multiple CIs in the analysis of factorial data. Adjustment of multiple comparisons is necessary.

The formulae for the Bonferroni, Scheffe, and Tukey methods can all be used similarly as was done in Chap.4. We will demonstrate the methods in the examples below.

```
popcorn$brand = as.factor(popcorn$brand)
popcorn$power = as.factor(popcorn$power)
popcorn$time = as.factor(popcorn$time)
lm1 = lm(y ~ brand*power*time, data=popcorn)
anova(lm1)
Analysis of Variance Table

Response: y
                 Df   Sum Sq Mean Sq F value    Pr(>F)
brand             2  331.101 165.550 1.88856 0.1800727
power             1  455.111 455.111 5.19181 0.0351175
time              2 1554.576 777.288 8.86713 0.0020878
brand:power       2  196.041  98.020 1.11819 0.3485423
brand:time        4 1433.858 358.464 4.08928 0.0157156
power:time        2   47.709  23.854 0.27213 0.7648363
brand:power:time  4   47.334  11.834 0.13500 0.9673241
Residuals        18 1577.870  87.659
```

We need MSE $= 87.659$ with df $= 18$ to calculate the SE's.

Which combination of brand, power, and time will produce the highest popping rate?

To answer this question, we can just conduct pairwise comparisons between all $3 \times 2 \times 3 = 18$ treatments. using Tukey's HSD, controlling FWER at α,

$$\text{HSD} = \frac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}\left(\frac{1}{r} + \frac{1}{r}\right)}.$$

Here $g = 18$ is the # of treatments, dfE $= 18$ is the df of SSE. At FWER $= 0.01$, Tukey's critical value is obtained below to be 4.846.

```
qtukey(1-0.01, 18, 18)/sqrt(2)
[1] 4.8462
```

As MSE $= 87.659$ (from the ANOVA table) and there are $n = 2$ replicates per treatment, Tukey's HSD is

$$4.846 \times \sqrt{87.659\left(\frac{1}{2} + \frac{1}{2}\right)} \approx 45.37.$$

```
library(mosaic)
sort(mean(y ~ brand+power+time, data=popcorn))
2.1.3 2.2.3 3.1.3 3.1.1 3.1.2 1.1.1 2.1.1 3.2.3 1.2.1 1.2.3
46.45 56.00 59.55 63.75 65.80 69.65 69.75 70.50 73.05 73.10
3.2.2 1.1.3 3.2.1 1.1.2 2.2.1 1.2.2 2.1.2 2.2.2
75.75 77.30 78.30 80.65 82.90 83.70 84.85 88.45
```

As maximum and minimum of the 18 group means differ by $88.45 - 46.45 = 42 < \text{HSD}$, no two groups are significantly different from each other, controlling FWER at 0.01, which is not surprising as the P-value for the overall treatment effect is 0.021.

```
anova(lm(y ~ brand:power:time, data=popcorn))
Analysis of Variance Table

Response: y
                 Df  Sum Sq  Mean Sq F value   Pr(>F)
brand:power:time 17 4065.73 239.1605 2.72829 0.020577
Residuals        18 1577.87  87.6594
```

Main-Effect Contrasts

Main-Effect Contrasts

When the coefficients c_{ijk} depend only on a single index, like $c_{ijk} = c_i$ depend only on index i for all i, j, k, then

$$C = \sum\nolimits_{ijk} c_{ijk}\mu_{ijk} = \sum\nolimits_{ijk} c_i\mu_{ijk} = \sum\nolimits_i c_i \sum\nolimits_{jk} \mu_{ijk} = \sum\nolimits_i c_i\mu_{i\bullet\bullet}$$

Recall $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$. Summing over indexes j and k, all other terms vanish except μ and α_i because of the zero-sum constraints,

$$\begin{aligned}\mu_{i\bullet\bullet} &= bc\mu + bc\alpha_i + c\beta_\bullet + b\gamma_\bullet + c\alpha\beta_{i\bullet} + b\alpha\gamma_{i\bullet} + \beta\gamma_{\bullet\bullet} + \alpha\beta\gamma_{i\bullet\bullet} \\ &= bc(\mu + \alpha_i).\end{aligned}$$

Such a contrast only depends on main effects for factor A.

$$C = \sum\nolimits_i c_i\mu_{i\bullet\bullet} = bc\underbrace{\sum\nolimits_i c_i}_{=0}\mu + bc\sum\nolimits_i c_i\alpha_i = bc\sum\nolimits_i c_i\alpha_i$$

To know which `brand` performs best overall, averaging over levels of `power` & `time`, the contrast $\bar{\mu}_{i_1\bullet\bullet} - \bar{\mu}_{i_2\bullet\bullet} = \alpha_{i_1} - \alpha_{i_2}$ is such a main effect contrast with

$$c_{ijk} = \begin{cases} \frac{1}{bc} & \text{if } i = i_1 \\ -\frac{1}{bc} & \text{if } i = i_2 \\ 0 & \text{if } i \neq i_1 \text{ or } i_2 \end{cases}$$

The SE for this contrast is $\mathrm{SE}(\widehat{C}) = \sqrt{\mathrm{MSE}}\sqrt{\sum_{ijk} \frac{c_{ijk}^2}{n_{ijk}}}$ where

$$\sum_{ijk} \frac{c_{ijk}^2}{n_{ijk}} = \sum_{jk} \frac{(1/bc)^2 + (-1/bc)^2}{n} = bc\frac{(1/bc)^2 + (-1/bc)^2}{n} = \frac{1}{nbc} + \frac{1}{nbc}.$$

Hence the SE for this contrast

$$\mathrm{SE}(\widehat{\bar{\mu}}_{i_1\bullet\bullet} - \widehat{\bar{\mu}}_{i_2\bullet\bullet}) = \mathrm{SE}(\bar{y}_{i_1\bullet\bullet\bullet} - \bar{y}_{i_2\bullet\bullet\bullet}) = \sqrt{\mathrm{MSE}\left(\frac{1}{nbc} + \frac{1}{nbc}\right)}$$

is just like the SE for pairwise comparisons of brands each with *nbc* replicates.

Which brand Works the Best Overall, averaging over power & time

Pairwise comparisons between all 3 brands: $\bar{\mu}_{1\bullet\bullet}, \bar{\mu}_{2\bullet\bullet}, \bar{\mu}_{3\bullet\bullet}$ using Tukey's HSD, controlling FWER at α,

$$\text{HSD} = \frac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}\left(\frac{1}{r} + \frac{1}{r}\right)}.$$

Here $g = 3$ is the # of brands compared, dfE $= 18$ is the df of SSE. Tukey's critical value at FWER $= 0.01$ is 3.3258 below.

```
qtukey(1-0.01, 3, 18)/sqrt(2)
[1] 3.32579
```

As MSE $= 87.659$ (from the ANOVA table) and there are $r = nbc = 2 \times 2 \times 3 = 12$ *replicates* per brand, Tukey's HSD is

$$3.326 \times \sqrt{87.659\left(\frac{1}{12} + \frac{1}{12}\right)} \approx 12.71.$$

```
library(mosaic)
sort(mean(y ~ brand, data=popcorn))
      3       2       1
68.9417 71.4000 76.2417
```

No two brands differ by more than HSD $= 12.71$ in mean at FWER $= 0.01$.

Not surprising since the P-value for the `brand` main effect is 0.18 in the ANOVA table.

```
library(emmeans)
lm1 = lm(y ~ brand*power*time, data=popcorn)
lm1embrand = emmeans(lm1, ~brand)
summary(contrast(lm1embrand, method="pairwise", adjust="tukey"),
        infer=c(T,T), level=0.99)
 contrast estimate   SE df lower.CL upper.CL t.ratio p.value
 1 - 2        4.84 3.82 18    -7.87     17.6   1.267  0.4312
 1 - 3        7.30 3.82 18    -5.41     20.0   1.910  0.1646
 2 - 3        2.46 3.82 18   -10.25     15.2   0.643  0.7985

Results are averaged over the levels of: power, time
Confidence level used: 0.99
Conf-level adjustment: tukey method for comparing a family of 3 estimat
P value adjustment: tukey method for comparing a family of 3 estimates
```

Which `time` Works the Best Overall, averaging over `brand` & `power`?

```
lm1emtime = emmeans(lm1, ~time)
summary(contrast(lm1emtime, method="pairwise", adjust="tukey"),
        infer=c(T,T), level=0.99)
 contrast estimate   SE df lower.CL upper.CL t.ratio p.value
 1 - 2       -6.97 3.82 18   -19.68     5.75  -1.823  0.1906
 1 - 3        9.08 3.82 18    -3.63    21.80   2.376  0.0705
 2 - 3       16.05 3.82 18     3.34    28.76   4.199  0.0015

Results are averaged over the levels of: brand, power
Confidence level used: 0.99
Conf-level adjustment: tukey method for comparing a family of 3 estimat
P value adjustment: tukey method for comparing a family of 3 estimates
```

Popping time = 2 (4.5 min) is significantly better than popping time = 3 (5 min) at FWER = 0.01.

Other pairs are not significantly different.

Which `power:time` Combination Works the Best, averaging over `brands`?

Could do pairwise comparisons $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ of all j_1, j_2, k_1, k_2 using Tukey's HSD, controlling FWER at α,

$$\text{HSD} = \frac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}\left(\frac{1}{r} + \frac{1}{r}\right)}.$$

Here $g = 2 \times 3 = 6$ is the # of `power`-`time` combinations, dfE $= 18$ is the df of SSE. At FWER $= 0.01$, Tukey's critical value is obtained below to be 3.9618

```
qtukey(1-0.01, 6, 18)/sqrt(2)
[1] 3.96177
```

As MSE $= 87.659$, and there are $r = na = 2 \times 3 = 6$ *replicates* per `power` $\times$ `time` combination, Tukey's HSD is

$$3.9618 \times \sqrt{87.659\left(\frac{1}{6} + \frac{1}{6}\right)} \approx 21.415.$$

```r
library(mosaic)
sort(mean(y ~ power+time, data=popcorn))
    1.3     2.3     1.1     1.2     2.1     2.2
61.1000 66.5333 67.7167 77.1000 78.0833 82.6333
```

Only (1,3)=(500W, 5 min) and (2,2) = (625W, 4.5 min) differ by more than HSD = 21.45 in mean at FWER = 0.01.

```
lm1empt = emmeans(lm1, ~power:time)
summary(contrast(lm1empt, method="pairwise", adjust="tukey"),
        infer=c(T,T), level=0.99)
 contrast  estimate   SE df lower.CL upper.CL t.ratio p.value
 1 1 - 2 1  -10.367 5.41 18  -31.782     11.0  -1.918  0.4235
 1 1 - 1 2   -9.383 5.41 18  -30.799     12.0  -1.736  0.5274
 1 1 - 2 2  -14.917 5.41 18  -36.332      6.5  -2.760  0.1111
 1 1 - 1 3    6.617 5.41 18  -14.799     28.0   1.224  0.8197
 1 1 - 2 3    1.183 5.41 18  -20.232     22.6   0.219  0.9999
 2 1 - 1 2    0.983 5.41 18  -20.432     22.4   0.182  1.0000
 2 1 - 2 2   -4.550 5.41 18  -25.965     16.9  -0.842  0.9554
 2 1 - 1 3   16.983 5.41 18   -4.432     38.4   3.142  0.0537
 2 1 - 2 3   11.550 5.41 18   -9.866     33.0   2.137  0.3131
 1 2 - 2 2   -5.533 5.41 18  -26.949     15.9  -1.024  0.9039
 1 2 - 1 3   16.000 5.41 18   -5.415     37.4   2.960  0.0764
 1 2 - 2 3   10.567 5.41 18  -10.849     32.0   1.955  0.4035
 2 2 - 1 3   21.533 5.41 18    0.118     42.9   3.984  0.0096
 2 2 - 2 3   16.100 5.41 18   -5.316     37.5   2.978  0.0737
 1 3 - 2 3   -5.433 5.41 18  -26.849     16.0  -1.005  0.9102

Results are averaged over the levels of: brand
Confidence level used: 0.99
Conf-level adjustment: tukey method for comparing a family of 6 estimat
P value adjustment: tukey method for comparing a family of 6 estimates
```

More on Factor Effects

In Lecture 15, we only showed how to compare levels of a factor averaging over levels of other factors, like

- brand effects average over levels of power and time: $\bar{\mu}_{1\bullet\bullet}, \bar{\mu}_{2\bullet\bullet}, \bar{\mu}_{3\bullet\bullet}$
- time effects average over levels of brand and power: $\bar{\mu}_{\bullet\bullet 1}, \bar{\mu}_{\bullet\bullet 2}, \bar{\mu}_{\bullet\bullet 3}$

Can we compare levels of a factor, but with other factors fixed at a certain level?

- brand effect when power = 625W, 4.5min Popping Time: $\mu_{122}, \mu_{222}, \mu_{322}$
- brand effect at 4.5min popping time, averaging over power: $\bar{\mu}_{1\bullet 2}, \bar{\mu}_{2\bullet 2}, \bar{\mu}_{3\bullet 2}$

Brand effect When power = 625W, 4.5min Popping Time

- pairwise comparisons of the 3 treatments: $\mu_{122}, \mu_{222}, \mu_{322}$
- Use Tukey's HSD $= \dfrac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}\left(\dfrac{1}{r} + \dfrac{1}{r}\right)}$
  - $g = 3$ is the # of treatments compared
  - dfE = 18 is the df of SSE
  - Tukey's critical value at FWER = 0.01 is 3.3258.

```
qtukey(1-0.01, 3, 18)/sqrt(2)
[1] 3.32579
```

As there are $r = 2$ replicates per treatment, Tukey's HSD is

$$3.3258 \times \sqrt{87.659\left(\frac{1}{2} + \frac{1}{2}\right)} \approx 31.14.$$

```r
sort(mean(y ~ brand, data=subset(popcorn, power == 2 & time == 2)))
    3     1     2
75.75 83.70 88.45
```

No two brands differ by > 1 HSD from each other.
$\Rightarrow$ No brands are significantly different from each other when poped 4.5mins using a 625W microwave.

Brand effect at 4.5min Popping Time, averaging over `power`

- pairwise comparisons between $\bar{\mu}_{1\bullet 2}, \bar{\mu}_{2\bullet 2}, \bar{\mu}_{3\bullet 2}$
- Use Tukey's HSD $= \dfrac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}\left(\dfrac{1}{r}+\dfrac{1}{r}\right)}$
  - $g = 3$ is the # of groups compared
  - dfE $= 18$ is the df of SSE
  - $r = nb = 2 \times 2 = 4$ replicates in each group compared
  - Tukey's critical value at FWER $= 0.01$ is 3.3258.

```
qtukey(1-0.01, 3, 18)/sqrt(2)
[1] 3.32579
```

$$\text{Tukey's HSD is } 3.3258 \times \sqrt{87.659\left(\frac{1}{4}+\frac{1}{4}\right)} \approx 22.02.$$

```
sort(mean(y ~ brand, data=subset(popcorn, time == 2)))
     3      1      2
70.775 82.175 86.650
```

Still, no brands are significantly different.

Interaction Contrasts

Interaction Contrasts

E.g., an AB interaction contrast is a contrast of the form

$$C = \sum_{ij} c_{ij}\bar{\mu}_{ij\bullet},$$

where the coefficients c_{ij} satisfy the zero-sum constraints $\sum_i c_{ij} = \sum_j c_{ij} = 0$ for all i, j.

As $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$, summing over index k, all terms involve index k vanish because of the zero-sum constraints,

$$\begin{aligned}\mu_{ij\bullet} &= c\mu + c\alpha_i + c\beta_j + \gamma_\bullet + c\alpha\beta_{ij} + \alpha\gamma_{i\bullet} + \beta\gamma_{j\bullet} + \alpha\beta\gamma_{ij\bullet} \\ &= c(\mu + \alpha_i + \beta_j + \alpha\beta_{ij}) \Rightarrow \bar{\mu}_{ij\bullet} = \frac{1}{c}\mu_{ij\bullet} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}\end{aligned}$$

Thus such a contrast compares the AB interactions.

$$\begin{aligned}C &= \sum_{ij} c_{ij}\bar{\mu}_{ij\bullet} = \sum_{ij} c_{ij}(\mu + \alpha_i + \beta_j + \alpha\beta_{ij}) \\ &= c_{\bullet\bullet}\mu + \sum_i c_{i\bullet}\alpha_i + \sum_j c_{\bullet j}\beta_j + \sum_{ij} c_{ij}\alpha\beta_{ij} = \sum_{ij} c_{ij}\alpha\beta_{ij}.\end{aligned}$$

Estimate and SE of an Interaction Contrast

The AB interaction

$$C = \sum_{ij} c_{ij}\bar{\mu}_{ij\bullet} = \sum_{ij} c_{ij}\alpha\beta_{ij}$$

can be estimated by

$$\widehat{C} = \sum_{ij} c_{ij}\bar{y}_{ij\bullet\bullet} = \sum_{ij} c_{ij}\widehat{\alpha\beta}_{ij}$$

where $\widehat{\alpha\beta}_{ij} = \bar{y}_{ij\bullet\bullet} - \bar{y}_{i\bullet\bullet\bullet} - \bar{y}_{\bullet j\bullet\bullet} + \bar{y}_{\bullet\bullet\bullet\bullet}$. The SE is

$$\mathrm{SE}(\widehat{C}) = \sqrt{\mathrm{MSE} \times \sum_{ij} \frac{c_{ij}^2}{r}}$$

where r is the number of observations used to calculate $\bar{y}_{ij\bullet\bullet}$, which $r = nc$ for $a \times b \times c$ designs with n replicates.

Scheffe's Adjustment for Interaction Contrasts

For an $a \times b \times c$ factorial design, to infer about the *family* of all AB interaction contrasts of the form

$$C = \sum_{ij} c_{ij}\bar{\mu}_{ij\bullet} = \sum_{ij} c_{ij}\alpha\beta_{ij}$$

- $100(1-\alpha)\%$ Scheffe's simultaneous C.I. for the contrast C is

$$\widehat{C} \pm \sqrt{df_{AB} F_{df_{AB}, dfE, \alpha}}\,\mathrm{SE}(\widehat{C})$$

- For testing $\mathrm{H}_0 : C = 0$ v.s. $\mathrm{H}_a : C \neq 0$, reject H_0 when

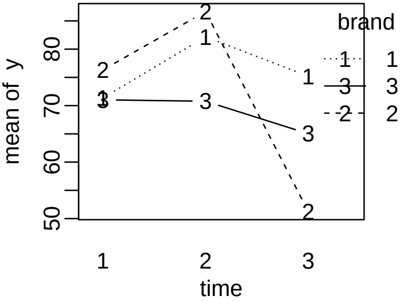
$$|t_0| = \frac{|\widehat{C}|}{\mathrm{SE}(\widehat{C})} > \sqrt{df_{AB} F_{df_{AB}, dfE, \alpha}}$$

  where

- $df_{AB} = (a-1)(b-1)$ is the df of AB interactions,
- dfE is the df of the SSE.

Example

From the brand-time interaction plot on the right, there seems to be interaction between brands 1 and 2 and popping times 2 and 3.



$$C = \bar{\mu}_{1\bullet 2} - \bar{\mu}_{1\bullet 3} - \bar{\mu}_{2\bullet 2} + \bar{\mu}_{2\bullet 3}$$
$$= \alpha\gamma_{12} - \alpha\gamma_{13} - \alpha\gamma_{22} + \alpha\gamma_{23},$$

can be estimated to be

$$\widehat{C} = \bar{y}_{1\bullet 2\bullet} - \bar{y}_{1\bullet 3\bullet} - \bar{y}_{2\bullet 2\bullet} + \bar{y}_{2\bullet 3\bullet}$$
$$= 82.175 - 75.2 - 86.65 + 51.225 = -28.45.$$

```
mean(y ~ brand+time, data=popcorn)
   1.1    2.1    3.1    1.2    2.2    3.2    1.3    2.3    3.3
71.350 76.325 71.025 82.175 86.650 70.775 75.200 51.225 65.025
```

The SE for the contrast $C = \bar{\mu}_{1\bullet 2} - \bar{\mu}_{1\bullet 3} - \bar{\mu}_{2\bullet 2} + \bar{\mu}_{2\bullet 3}$ is

$$\begin{aligned}\mathrm{SE}(\widehat{C}) &= \sqrt{\mathrm{MSE} \times \sum_{ij} \frac{c_{ij}^2}{nb}} \\ &= \sqrt{87.659\left(\frac{1^2+(-1)^2+(-1)^2+1^2}{2\times 2}\right)} \approx 9.363\end{aligned}$$

where the denominator is $nb = (2)(2) = 4$ as `power` has 2 levels.

The t-statistic for testing H_0: $C = \bar{\mu}_{1\bullet 2} - \bar{\mu}_{1\bullet 3} - \bar{\mu}_{2\bullet 2} + \bar{\mu}_{2\bullet 3} = 0$ is

$$t = \frac{\widehat{C}}{\mathrm{SE}(\widehat{C})} \approx \frac{-28.45}{9.363} \approx -3.04 \quad \text{w/ 18 df.}$$

Not adjusted multiple comparisons, the 2-sided P-value is ≈ 0.007.

```
2*pt(-3.04,df=18)
[1] 0.00704439
```

Can we conclude this interaction contrast is significantly different from 0 as the P-value is below 0.01?

Scheffe's Adjustment for Interaction Contrasts

- As we decided to test this contrast after looking at the `brand-time` interaction plot (data snooping), should use Scheffe's adjustment
- For the *family* of AC interactions, Scheffe's critical value is $\sqrt{df_{AC} F_{df_{AC}, dfE, \alpha}}$ at FWER $= \alpha$.
  - Here $df_{AC} = (a-1)(c-1) = (3-1)(3-1) = 4$ as `brand` and `time` both have 3 levels
  - dfE $= 18$ is the df of SSE

```
sqrt((3-1)*(3-1)*qf(0.05, (3-1)*(3-1), 18, lower.tail=FALSE))
[1] 3.42213
```

At FWER $= 0.05$, Scheffe's critical value is ≈ 3.422 above, which is over the $|t - statistic| = 3.04$ for the contrast. The contrast is hence not significant at FWER $= 0.05$.

```
lm1embt = emmeans(lm1, ~brand:time)
summary(contrast(lm1embt, list(C=c(0,0,0,1,-1,0,-1,1,0)),
                 infer=c(T,T),adjust="scheffe"),scheffe.rank=4)
 contrast estimate   SE df lower.CL upper.CL t.ratio p.value
 C           -28.4 9.36 18    -60.5     3.59  -3.039  0.0975

Results are averaged over the levels of: power
Confidence level used: 0.95
Conf-level adjustment: scheffe method with rank 4
P value adjustment: scheffe method with rank 4
```

Note `scheffe.rank=4` since $df_{AC} = (3-1)(3-1) = 4$.

Observe 95% Scheffe's CI for the contrast is

$$\widehat{C} \pm (\text{crit. val})\text{SE} \approx -28.45 \pm 3.422 \times 9.363 \approx (-60.49, 3.59).$$

which is exactly the CI (-60.5,3.59) given by the `emmeans` library.

Bonferroni's Method

Suppose we just focused on AC interaction contrasts of the form

$$C = \bar{\mu}_{i_1 \bullet k_1} - \bar{\mu}_{i_1 \bullet k_2} - \bar{\mu}_{i_2 \bullet k_1} + \bar{\mu}_{i_2 \bullet k_2}$$

that check the interactions of levels (i_1, i_2) of factor A and levels (k_1, k_2) of factor C, rather than contrasts with arbitrary c_{ij}'s,

- Effectively there are only $\binom{a}{2}\binom{c}{2}$ such contrasts, as there are $\binom{a}{2}$ ways to choose 2 levels from the a levels of Factor A, and $\binom{c}{2}$ ways to choose 2 levels from the c levels of Factor C

For such a family of contrasts, we can control the FWER by Bonferroni's method

- $100(1-\alpha)\%$ Bonferro's simultaneous C.I. for the contrast C is

$$\widehat{C} \pm t_{dfE,\alpha/2/m}\mathrm{SE}(\widehat{C})$$

- For testing $\mathrm{H}_0 : C = 0$ v.s. $\mathrm{H}_a : C \neq 0$, reject H_0 when

$$|t_0| = \frac{|\widehat{C}|}{\mathrm{SE}(\widehat{C})} > t_{dfE,\alpha/2/m}$$

where dfE is the df of the SSE, and $m = \binom{a}{2}\binom{c}{2}$.

For the Popcorn Data, to see if any of the brand-time interaction contrast of the form

$$C = \bar{\mu}_{i_1 \bullet k_1} - \bar{\mu}_{i_1 \bullet k_2} - \bar{\mu}_{i_2 \bullet k_1} + \bar{\mu}_{i_2 \bullet k_2}$$

is significantly different from 0, Bonferroni's critical value controlling FWER at 0.05 is 3.149.

```
qt(0.05/2/9, df=18, lower.tail=FALSE)
[1] 3.14858
```

Here $m = \binom{a}{2}\binom{c}{2} = \binom{3}{2}\binom{3}{2} = 3 \times 3 = 9$ because both brand and time have 3 levels.

For the contrast $C = \bar{\mu}_{1\bullet 2} - \bar{\mu}_{1\bullet 3} - \bar{\mu}_{2\bullet 2} + \bar{\mu}_{2\bullet 3}$, it's t-statistic in absolute value 3.04 is below 3.149 and hence it's not significantly different from 0.

Scheffe's Adjustment for All Contrasts

Suppose we checked the all the 2-way and 3-way interaction plots and noticed the contrast below is most likely to be significant

$$C = \bar{\mu}_{1\bullet 2} - \bar{\mu}_{1\bullet 3} - \bar{\mu}_{2\bullet 2} + \bar{\mu}_{2\bullet 3}.$$

Then effectively, we have considered all the possible contrasts and hence must use Scheffe's adjustment for *ALL contrasts*, where the critical value at FWER $= \alpha$ is

$$\sqrt{(g-1)F_{g-1,dfE,\alpha}}$$

For the Popcorn data, $g = abc = (3)(2)(3) = 18$ since we have considered all the possible interactions between the 3 factors, df $=$ 18 is the df of SSE.

At FWER $= 0.05$, Scheffe's critical value is about 6.161.

```
sqrt((18-1)*qf(0.05, 18-1, 18, lower.tail=FALSE))
[1] 6.16062
```

The contrast is NOT significant at FWER $= 0.01$ since $|\text{t-statistic}| = 3.04$ is below Scheffe's critical value 6.161.

Summary of Multiple Comparisons for Intraction Contrasts

- For the family of all `brand`-`time` interaction contrasts

$$C = \sum_{ik} c_{ik}\bar{\mu}_{i\bullet k} \quad \text{with } \sum_i c_{ik} = \sum_k c_{ij} = 0, \text{ for all } i, k,$$

  Scheffe's critical value at FWER $= \alpha = 0.05$ is

$$\sqrt{df_{AC} F_{df_{AC}, dfE, \alpha}} = 3.42, \quad \text{where } df_{AC} = (a-1)(c-1).$$

- For the family of all `brand`-`time` interaction contrasts of the form $C = \bar{\mu}_{i_1\bullet k_1} - \bar{\mu}_{i_1\bullet k_2} - \bar{\mu}_{i_2\bullet k_1} + \bar{\mu}_{i_2\bullet k_2}$, Bonferroni's critical value at FWER $= \alpha = 0.05$ is

$$t_{dfE, \alpha/2/m} = 3.149, \quad \text{where } m = \binom{a}{2}\binom{c}{2} = \binom{3}{2}\binom{3}{2} = 3 \times 3 = 9$$

- For the family of ALL possible contrasts of the $g = 18$ treatments, Scheffe's critical value at FWER $= \alpha = 0.05$ is

$$\sqrt{(g-1)F_{g-1, dfE, \alpha}} = 6.161.$$

Summary of Multiple Comparisons for Intraction Contrasts

From the 3 critical values on the previous page, we can see

the larger the family of tests, the greater the critical value

It's important to *outline contrasts of interest (pre-planned contrasts)* before looking at data.

If a contrast is not pre-planned, the critical value for it to be significant can be much bigger.

Controlling the FWER for the Entire Analysis

FWER for the Entire Analysis of the Popcorn Data

So far, we have conducted the following tests for the Popcorn data.

1. pairwise comparisons of all 18 treatments: $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. pairwise comparisons of all 3 `brand`s, averaging over levels of `power` & `time`: $\bar{\mu}_{i_1 \bullet\bullet} - \bar{\mu}_{i_2 \bullet\bullet}$
3. pairwise comparisons of the 3 `brand`s when `power`$=$ 625W, `time` $=$ 4.5min: $\mu_{122}, \mu_{222}, \mu_{322}$
4. pairwise comparisons of the 3 `brand`s when `time` $=$ 4.5min, averaging over `power`: $\bar{\mu}_{1\bullet 2}, \bar{\mu}_{2\bullet 2}, \bar{\mu}_{3\bullet 2}$
5. pairwise comparisons of the 3 popping `time`s, averaging over levels of `brand` & `power`: $\bar{\mu}_{\bullet\bullet 1}, \bar{\mu}_{\bullet\bullet 2}, \bar{\mu}_{\bullet\bullet 3}$
6. pairwise comparisons of the 6 combinations of `power` & `time`, averaging over the 3 `brand`s: $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ for all j_1, j_2, k_1, k_2
7. interaction contrasts of `brand` & `time`

The first six families are tested at FWER $= 0.01$, and the 7th at FWER $= 0.05$. By Bonferroni's inequality, the overall FWER for all 7 families is at most

FWER for the Entire Analysis of the Popcorn Data

So far, we have conducted the following tests for the Popcorn data.

1. pairwise comparisons of all 18 treatments: $\mu_{i_1 j_1 k_1} - \mu_{i_2 j_2 k_2}$ of all $i_1, i_2, j_1, j_2, k_1, k_2$
2. pairwise comparisons of all 3 `brand`s, averaging over levels of `power` & `time`: $\bar{\mu}_{i_1 \bullet\bullet} - \bar{\mu}_{i_2 \bullet\bullet}$
3. pairwise comparisons of the 3 `brand`s when `power`$=$ 625W, `time` $=$ 4.5min: $\mu_{122}, \mu_{222}, \mu_{322}$
4. pairwise comparisons of the 3 `brand`s when `time` $=$ 4.5min, averaging over `power`: $\bar{\mu}_{1\bullet 2}, \bar{\mu}_{2\bullet 2}, \bar{\mu}_{3\bullet 2}$
5. pairwise comparisons of the 3 popping `time`s, averaging over levels of `brand` & `power`: $\bar{\mu}_{\bullet\bullet 1}, \bar{\mu}_{\bullet\bullet 2}, \bar{\mu}_{\bullet\bullet 3}$
6. pairwise comparisons of the 6 combinations of `power` & `time`, averaging over the 3 `brand`s: $\bar{\mu}_{\bullet j_1 k_1} - \bar{\mu}_{\bullet j_2 k_2}$ for all j_1, j_2, k_1, k_2
7. interaction contrasts of `brand` & `time`

The first six families are tested at FWER $= 0.01$, and the 7th at FWER $= 0.05$. By Bonferroni's inequality, the overall FWER for all 7 families is at most $0.01 + 0.01 + 0.01 + 0.01 + 0.01 + 0.01 + 0.05 = 0.11$.

Controlling the FWER for the Entire Analysis

In general, prior to looking at the data, one should outline of contrasts of interest. For factorial data, these may include

- pairwise comparisons of main effects for each factor
- pairwise comparisons of combinations of two or more factors
- interaction contrasts of two factors
- (interaction contrasts of two factors)

For each family outlined, we may control the FWER for the family using Tukey's, Scheffe's or Bonferroni's methods, depending on the type.

Then the FWER for the entire analysis is no more than the sum of FWER for each family by Bonferroni's inequality.

Example: Controlling the FWER for Two-Way Data

For two-way data, we might be interested in

- comparison of all treatments (AB combinations),
- comparison of main effects of A, and
- comparison of main effects of B.

We may set the FWER for the 3 sets of inferences to be α_1, α_2, and α_3.

Then, by Bonferroni's inequality, the overall FWER is at most $\alpha_1 + \alpha_2 + \alpha_3$.

Section 6.5.6
Model Building

Section 6.5.6 Model Building

In some experiments, the primary objective is to find a model that gives an adequate representation for data, called *model building*.

For full k-way factorial data, it is *legitimate* to

- begin with the full model with all main effects and 2-way, 3-way, ... and k-way interactions,
- test the significance of the main effects and interactions, and
- adopt the smallest *hierarchical* model that includes all the significant terms as a reasonable model to represent the same type of experimental data in *future* experiments.

Note that it is *NOT legitimate* to adopt the simplified model and to use the corresponding ANOVA table, to test further hypotheses or calculate confidence intervals using *the same set of data*.

- If this is done, the model is changed based on the data, and the quoted significance levels and confidence levels associated with further inferences will not be correct.
- The data used for **Model building** should be **completely different** from the data used for further analysis based on the selected model.

Example: Popcorn Data

The analysis conducted for the Popcorn data in L15-16 are all based on the assumption that the full 3-way model is correct.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
brand	2	331	166	1.89	0.1801
power	1	455	455	5.19	0.0351
time	2	1555	777	8.87	0.0021
brand:power	2	196	98	1.12	0.3485
brand:time	4	1434	358	4.09	0.0157
power:time	2	48	24	0.27	0.7648
brand:power:time	4	47	12	0.13	0.9673
Residuals	18	1578	88		

As several main effects and interactions have large P-values, can we based our analysis of contrasts on the simplified model below?

```
lm2 = lm(y ~ brand+power+time+brand*time, data=popcorn)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
brand	2	331	166	2.30	0.11999
power	1	455	455	6.33	0.01837
time	2	1555	777	10.81	0.00038
brand:time	4	1434	358	4.99	0.00405
Residuals	26	1869	72		

Example: Popcorn Data

- It might be justifiable if the new model is applied on new data from experiments of the same type
  - Still risky to drop the insignificant terms since the main effects or interactions might be insignificant because of the small sample size. Insignificance doesn't prove they don't exist.
- The answer is *No* if applying the new model on the original data
  - The terms that are omitted from the model are those with small SS (Sum of Squares).
  - The df and SS of those omitted terms are pooled into the df and SS of errors in the simplified model. Hence the MSE of the new model could underestimate σ^2, and is no longer an unbiased estimate for σ^2, affect the validity of all the analysis that used the model.

If one has strong belief (based on past studies etc) that some interactions don't exist, it's okay to analyze data based on some simpler model rather than the full k-way model.

- Such models are used for
  - "single-replicate" data
  - confounded block-factorial design (chapter 13)
  - fractional factorial design (chapter 15)
- The model must be determined before looking at the data, or even before the experiment is conducted
- dfE is the df of SSE for the model used