STAT 222 Lecture 13-14
Chapter 7 General Factorial Designs

Yibi Huang

Outline

- General Factorial Designs
  - Definition of 3-way and k-way interactions
  - 3-way interaction plots
  - Parameter estimates
  - Sum of Squares, dfs, and the ANOVA table
- Hierarchy

3-Way and k-Way Interactions

3-Way Interaction Contrast

Based on the means model $y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ijk\ell}$ of a 3-way design, a *3-way interaction contrast* between level (i_1, i_2) for factor A, level (j_1, j_2) for factor B, and level (k_1, k_2) for factor C is defined to be

$$\mu_{i_1 j_1 k_1} - \mu_{i_2 j_1 k_1} - \mu_{i_1 j_2 k_1} - \mu_{i_1 j_1 k_2} + \mu_{i_2 j_2 k_1} + \mu_{i_2 j_1 k_2} + \mu_{i_1 j_2 k_2} - \mu_{i_2 j_2 k_2}$$

Observe that any two μ_{ijk}'s in the contrast have

opposite
identical signs if they differ by an **odd**
even number of indexes.

3-Way Interaction Contrast

Based on the means model $y_{ijk\ell} = \mu_{ijk} + \varepsilon_{ijk\ell}$ of a 3-way design, a *3-way interaction contrast* between level (i_1, i_2) for factor A, level (j_1, j_2) for factor B, and level (k_1, k_2) for factor C is defined to be

$$\mu_{i_1j_1k_1} - \mu_{i_2j_1k_1} - \mu_{i_1j_2k_1} - \mu_{i_1j_1k_2} + \mu_{i_2j_2k_1} + \mu_{i_2j_1k_2} + \mu_{i_1j_2k_2} - \mu_{i_2j_2k_2}$$

Observe that any two μ_{ijk}'s in the contrast have

opposite / **identical** signs if they differ by an **odd** / **even** number of indexes.

The 3-way interaction contrast above has 3 interpretations:

$$\begin{aligned}
&\mu_{i_1j_1k_1} - \mu_{i_2j_1k_1} - \mu_{i_1j_2k_1} - \mu_{i_1j_1k_2} + \mu_{i_2j_2k_1} + \mu_{i_2j_1k_2} + \mu_{i_1j_2k_2} - \mu_{i_2j_2k_2} \\
&= \big(\underbrace{\mu_{i_1j_1k_1} - \mu_{i_2j_1k_1} - \mu_{i_1j_2k_1} + \mu_{i_2j_2k_1}}_{\text{AB interaction contrast when } C = k_1}\big) - \big(\underbrace{\mu_{i_1j_1k_2} - \mu_{i_2j_1k_2} - \mu_{i_1j_2k_2} + \mu_{i_2j_2k_2}}_{\text{AB interaction contrast when } C = k_2}\big) \\
&= \big(\underbrace{\mu_{i_1j_1k_1} - \mu_{i_1j_2k_1} - \mu_{i_1j_1k_2} + \mu_{i_1j_2k_2}}_{\text{BC interaction contrast when } A = i_1}\big) - \big(\underbrace{\mu_{i_2j_1k_1} - \mu_{i_2j_2k_1} - \mu_{i_2j_1k_2} + \mu_{i_2j_2k_2}}_{\text{BC interaction contrast when } A = i_2}\big) \\
&= \big(\underbrace{\mu_{i_1j_1k_1} - \mu_{i_2j_1k_1} - \mu_{i_1j_1k_2} + \mu_{i_2j_1k_2}}_{\text{AC interaction contrast when } B = j_1}\big) - \big(\underbrace{\mu_{i_1j_2k_1} - \mu_{i_2j_2k_1} - \mu_{i_1j_2k_2} + \mu_{i_2j_2k_2}}_{\text{AC interaction contrast when } B = j_2}\big)
\end{aligned}$$

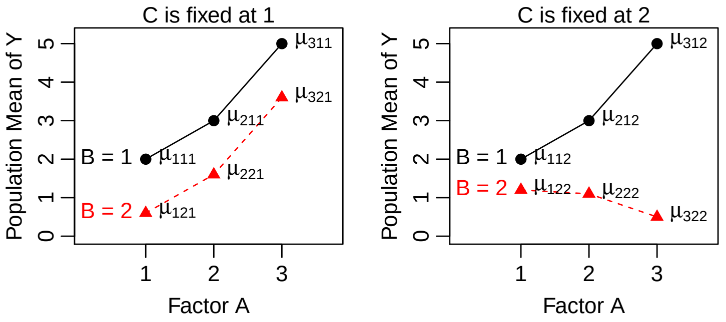
3-Way Interaction Plots

Three-Way Interactions

We say factors A, B, and C have *three-way interactions* if

- an AB interaction contrast changes with the levels of C, or
- a BC interaction contrast changes with the levels of A, or
- an AC interaction contrast changes with the levels of B.

E.g.,

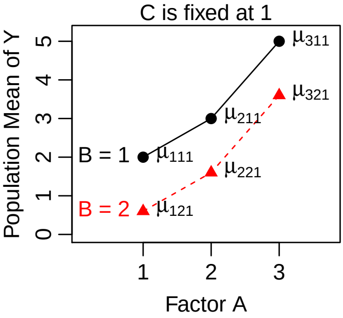


Three-Way Interactions

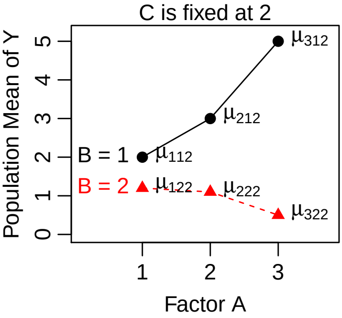
We say factors A, B, and C have *three-way interactions* if

- an AB interaction contrast changes with the levels of C, or
- a BC interaction contrast changes with the levels of A, or
- an AC interaction contrast changes with the levels of B.

E.g.,

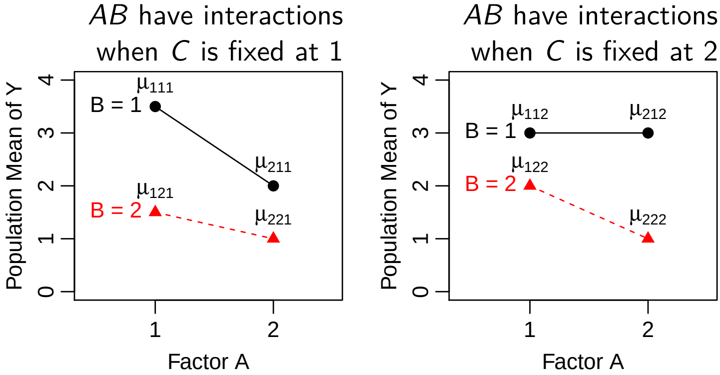


No AB interactions when C is fixed at 1

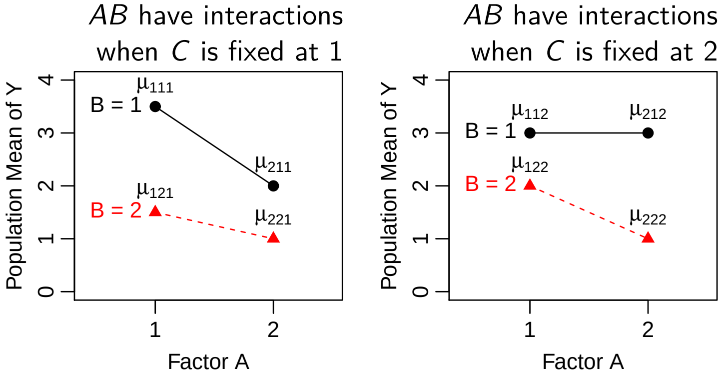


AB have interactions when C is fixed at 2

Example 2: Three-Way Interactions



Example 2: Three-Way Interactions

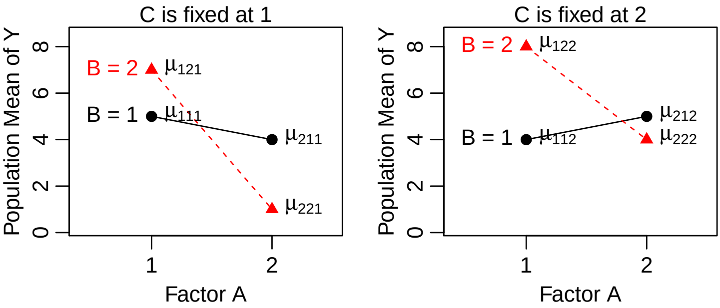


$$(\mu_{111} - \mu_{211}) - (\mu_{121} - \mu_{221}) > 0$$

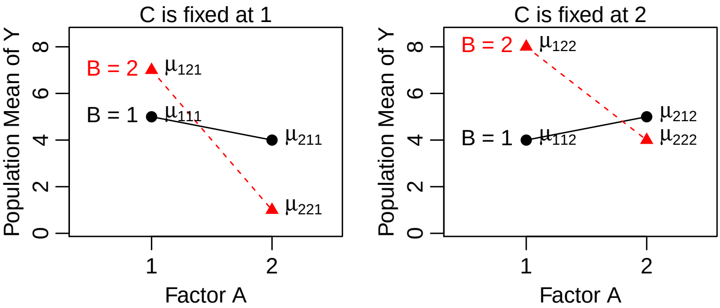
$$\underbrace{(\mu_{112} - \mu_{212})}_{=0} - \underbrace{(\mu_{122} - \mu_{222})}_{>0} < 0$$

The AB interaction contrast $(\mu_{11k} - \mu_{21k}) - (\mu_{12k} - \mu_{22k})$ depends on the level k of factor C. Hence there exist ABC 3-way interactions.

It can be hard to tell graphically whether ABC interaction is present when AB interactions exist at both levels of C.



It can be hard to tell graphically whether ABC interaction is present when AB interactions exist at both levels of C.



	$C=1$		$C=2$	
	$B=1$	$B=2$	$B=1$	$B=2$
$A=1$	$\mu_{111}=5$	$\mu_{121}=7$	$\mu_{112}=4$	$\mu_{122}=8$
$A=2$	$\mu_{211}=4$	$\mu_{221}=1$	$\mu_{212}=5$	$\mu_{222}=4$

The AB interactions at the two levels of C are equal and hence there is no ABC interaction.

$$(\mu_{111}-\mu_{211})-(\mu_{121}-\mu_{221})=(5-4)-(7-1)=-5$$
$$(\mu_{112}-\mu_{212})-(\mu_{122}-\mu_{222})=(4-5)-(8-4)=-5$$

Higher Order Interactions

- An ABCD 4-way interaction contrast is
  - the difference of some ABC 3-way interaction contrast at two different levels of D
  - the difference of some ABD 3-way interaction contrast at two different levels of C
  - the difference of some ACD 3-way interaction contrast at two different levels of B
  - the difference of some BCD 3-way interaction contrast at two different levels of A

Higher Order Interactions

- An ABCD 4-way interaction contrast is
  - the difference of some ABC 3-way interaction contrast at two different levels of D
  - the difference of some ABD 3-way interaction contrast at two different levels of C
  - the difference of some ACD 3-way interaction contrast at two different levels of B
  - the difference of some BCD 3-way interaction contrast at two different levels of A
- We say ABCD have *4-way interactions* if any of the ABCD 4-way interaction contrast is non-zero or if any 3-way interaction contrast between any 3 of the 4 factors changes with the levels of a 4th factor.
  - e.g., if some ACD 3-way interaction contrast changes with the levels of factor B, then there exist ABCD 4-way interaction
  - We say k factors have *k-way interactions* means the $(k-1)$-way interaction of any $(k-1)$ of the k factors changes with the levels of a kth factor.

General Factorial Models

General Factorial Models

The model and analysis of multi-way factorial data are generalization of those for two-way factorial data. E.g., consider a 4-way factorial design with factors A, B, C, and D.

$$\text{means model}: y_{ijk\ell m} = \mu_{ijk\ell} + \varepsilon_{ijk\ell m} \quad \text{for} \begin{cases} i = 1, \dots, a, \ j = 1, \dots, b, \\ k = 1, \dots, c, \ \ell = 1, \dots, d, \\ m = 1, \dots, n. \end{cases}$$

$$\begin{aligned} \text{effects model: } y_{ijk\ell m} = & \underbrace{\mu}_{\text{grand mean}} + \underbrace{\alpha_i + \beta_j + \gamma_k + \delta_\ell}_{\text{main effects}} \\ & + \underbrace{\alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\delta_{i\ell} + \beta\gamma_{jk} + \beta\delta_{j\ell} + \gamma\delta_{k\ell}}_{\text{2-way interactions}} \\ & + \underbrace{\alpha\beta\gamma_{ijk} + \alpha\beta\delta_{ij\ell} + \alpha\gamma\delta_{ik\ell} + \beta\gamma\delta_{jk\ell}}_{\text{3-way interactions}} \\ & + \underbrace{\alpha\beta\gamma\delta_{ijk\ell}}_{\text{4-way interaction}} + \underbrace{\varepsilon_{ijk\ell m}}_{\text{error}} \end{aligned}$$

Zero-Sum Constraints for General Factorial Models

$$\begin{aligned} y_{ijk\ell m} = \mu &+ \alpha_i + \beta_j + \gamma_k + \delta_\ell \\ &+ \alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\delta_{i\ell} + \beta\gamma_{jk} + \beta\delta_{j\ell} + \gamma\delta_{k\ell} \\ &+ \alpha\beta\gamma_{ijk} + \alpha\beta\delta_{ij\ell} + \alpha\gamma\delta_{ik\ell} + \beta\gamma\delta_{jk\ell} \\ &+ \alpha\beta\gamma\delta_{ijk\ell} + \varepsilon_{ijk\ell m} \end{aligned}$$

All the effects have zero-sum constraints that they add to 0 when summing over any subscript, e.g.,

- $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \sum_\ell \delta_\ell = 0$
- $\sum_i \alpha\gamma_{ik} = \sum_k \alpha\gamma_{ik} = 0$, for all i, k, so do other 2-way interactions
- $\sum_i \alpha\gamma\delta_{ik\ell} = \sum_k \alpha\gamma\delta_{ik\ell} = \sum_\ell \alpha\gamma\delta_{ik\ell} = 0$, for all i, k, ℓ, so do other 3-way interactions
- $\sum_i \alpha\beta\gamma\delta_{ijk\ell} = \sum_j \alpha\beta\gamma\delta_{ijk\ell} = \sum_k \alpha\beta\gamma\delta_{ijk\ell} = \sum_\ell \alpha\beta\gamma\delta_{ijk\ell} = 0$, for all i, j, k, ℓ.

Parameter Estimates

Parameter Estimates

For a 4-way model, the parameter estimates under the zero-sum constraints are

grand mean	$\widehat{\mu} = \bar{y}_{\bullet\bullet\bullet\bullet\bullet}$
main effects	$\widehat{\alpha}_i = \bar{y}_{i\bullet\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet}, \quad \widehat{\beta}_j = \bar{y}_{\bullet j\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet},$ $\widehat{\gamma}_k = \bar{y}_{\bullet\bullet k\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet}, \quad \widehat{\delta}_\ell = \bar{y}_{\bullet\bullet\bullet\ell\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet}$
2-way	$\widehat{\alpha\beta}_{ij} = \bar{y}_{ij\bullet\bullet\bullet} - \bar{y}_{i\bullet\bullet\bullet\bullet} - \bar{y}_{\bullet j\bullet\bullet\bullet} + \bar{y}_{\bullet\bullet\bullet\bullet\bullet}$ $\widehat{\beta\gamma}_{jk} = \bar{y}_{\bullet jk\bullet\bullet} - \bar{y}_{\bullet j\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet k\bullet\bullet} + \bar{y}_{\bullet\bullet\bullet\bullet\bullet}$ $\vdots$
3-way	$\widehat{\alpha\beta\delta}_{ij\ell} = \bar{y}_{ij\bullet\ell\bullet} - \bar{y}_{ij\bullet\bullet\bullet} - \bar{y}_{i\bullet\bullet\ell\bullet} - \bar{y}_{\bullet j\bullet\ell\bullet}$ $+\bar{y}_{i\bullet\bullet\bullet\bullet} + \bar{y}_{\bullet j\bullet\bullet\bullet} + \bar{y}_{\bullet\bullet\bullet\ell\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet}$ $\widehat{\alpha\gamma\delta}_{ik\ell} = \cdots$
4-way	$\widehat{\alpha\beta\gamma\delta}_{ijk\ell} =$ (16 terms, see the next page)

$$
\begin{aligned}
\widehat{\alpha\beta\gamma\delta}_{ijk\ell} = {} & \bar{y}_{ijk\ell\bullet} \\
& - \bar{y}_{ijk\bullet\bullet} - \bar{y}_{ij\bullet\ell\bullet} - \bar{y}_{i\bullet k\ell\bullet} - \bar{y}_{\bullet jk\ell\bullet} \\
& + \bar{y}_{ij\bullet\bullet\bullet} + \bar{y}_{i\bullet k\bullet\bullet} + \bar{y}_{i\bullet\bullet\ell\bullet} + \bar{y}_{\bullet jk\bullet\bullet} + \bar{y}_{\bullet j\bullet\ell\bullet} + \bar{y}_{\bullet\bullet k\ell\bullet} \\
& - \bar{y}_{i\bullet\bullet\bullet\bullet} - \bar{y}_{\bullet j\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet k\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\ell\bullet} \\
& + \bar{y}_{\bullet\bullet\bullet\bullet\bullet} \\
= {} & (\text{terms that average over 1 index}) \\
& - (\text{terms that average over 2 indexes}) \\
& + (\text{terms that average over 3 indexes}) \\
& - (\text{terms that average over 4 indexes}) \\
& + (\text{terms that average over 5 indexes})
\end{aligned}
$$

Sum of Squares

Sum of Squares

SST can be decomposed into SS of main effects and interactions of all orders, e.g., in an $a \times b \times c \times d$ design with n replicates:

$$\begin{aligned} SST = & SS_A + SS_B + SS_C + SS_D \\ & + SS_{AB} + SS_{AC} + SS_{AD} + SS_{BC} + SS_{BD} + SS_{CD} \\ & + SS_{ABC} + SS_{ACD} + SS_{ABD} + SS_{BCD} \\ & + SS_{ABCD} \\ & + SSE \end{aligned}$$

where SST $= \sum_{ijk\ell m}(y_{ijk\ell m} - \bar{y}_{\bullet\bullet\bullet\bullet\bullet})^2$, SSE $= \sum_{ijk\ell m}(y_{ijk\ell m} - \bar{y}_{ijk\ell\bullet})^2$, and the SS for all other terms are the **sum of squares of corresponding parameter estimates <u>under the zero sum constraints</u>**, e.g.,

$$\begin{aligned} SS_C &= \sum_{ijk\ell m}(\widehat{\gamma}_k)^2 = abdn \sum_k (\widehat{\gamma}_k)^2 \\ SS_{BC} &= \sum_{ijk\ell m}(\widehat{\beta\gamma}_{jk})^2 = adn \sum_{jk} (\widehat{\beta\gamma}_{jk})^2 \\ SS_{ACD} &= \sum_{ijk\ell m}(\widehat{\alpha\gamma\delta}_{ik\ell})^2 = bn \sum_{ik\ell} (\widehat{\alpha\gamma\delta}_{ik\ell})^2 \\ SS_{ABCD} &= \sum_{ijk\ell m}(\widehat{\alpha\beta\gamma\delta}_{ijk\ell})^2 = n \sum_{ijk\ell} (\widehat{\alpha\beta\gamma\delta}_{ijk\ell})^2 \end{aligned}$$

Degrees of Freedom

Say factors A, B, C, and D have respectively a, b, c, and d levels, and there are n replicates.

- d.f. of a main effect = number of levels -1.
  e.g., $df_A = a - 1$, $df_C = c - 1$.
- d.f. of an interaction = product of d.f.'s for the main effects of the involved factors, e.g.,
  - $df_{AD} = (a-1)(d-1)$,
  - $df_{BCD} = (b-1)(c-1)(d-1)$,
  - $df_{ABCD} = (a-1)(b-1)(c-1)(d-1)$.
- d.f. of SST = total # of observation $-1 = abcdn - 1$
- d.f. of SSE = total # of observation $-$ total # of treatments
  $= abcdn - abcd = abcd(n-1)$

Example: Popcorn Microwave Data (Section 7.4)

- A $3 \times 2 \times 3$ factorial design with 3 factors:
  - `brand`: 3 brands of popcorn, labelled 1, 2, 3
  - `power`: power of the microwave oven (1 = 500W, 2 = 625W)
  - `time`: popping time (1 = 4 mins, 2 = 4.5 mins, 3 = 5 mins)
- 2 replicates per treatment
- Response: % of kernels popped successfully in a package
- Must read Section 7.4 for study design details

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

Loading data:

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

Need to convert the 3 variables to factors before fit the model.

```
popcorn$brand = as.factor(popcorn$brand)
popcorn$power = as.factor(popcorn$power)
popcorn$time = as.factor(popcorn$time)
```

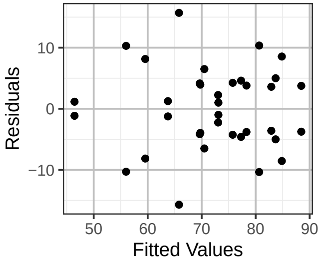
Model:

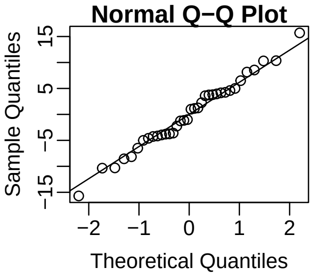
```
lm1 = lm(y ~ brand*power*time, data=popcorn)
```

Always check model assumptions FIRST!

Checking Model Assumptions — Popcorn Data

```
library(ggplot2)
ggplot(popcorn, aes(x=fitted(lm1), y=lm1$res))+geom_point()+
  labs(x="Fitted Values", y="Residuals")
qqnorm(lm1$res)
qqline(lm1$res)
```

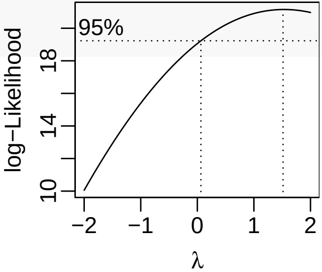




- Size of residuals doesn't seem to increase or decrease w/ fitted values
- Residuals appear normally distributed

Box-Cox — Popcorn Data

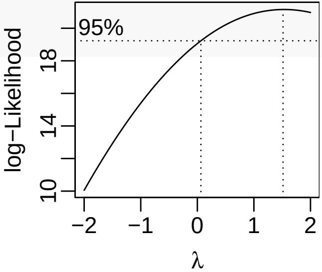
```
library(MASS)
boxcox(lm1)
```



- Only *two*, not 3, vertical dotted lines in the Box-Cox plot. Why?

Box-Cox — Popcorn Data

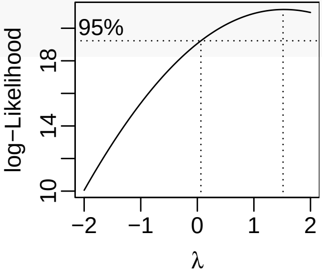
```
library(MASS)
boxcox(lm1)
```



- Only *two*, not 3, vertical dotted lines in the Box-Cox plot. Why?
  - The 3rd line is outside of the plot (not between -2 and 2)

Box-Cox — Popcorn Data

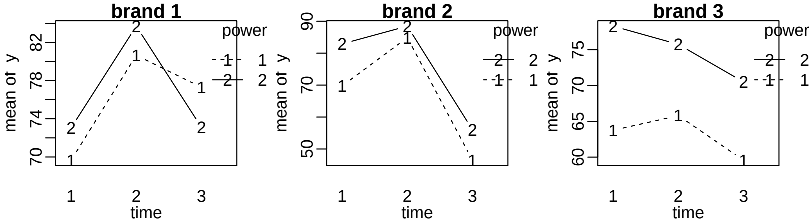
```
library(MASS)
boxcox(lm1)
```



- Only *two*, not 3, vertical dotted lines in the Box-Cox plot. Why?
  - The 3rd line is outside of the plot (not between -2 and 2)
- The line around $\lambda = 1.5$ is the optimal λ since it's at the maximum of the curve. The line around $\lambda = 0$ thus must be the lower bound of the 95% CI and the upper bound is over 2.
- 95% CI for λ includes 1, $\Rightarrow$ no transformation is required.

3-Way Interaction Plots — Popcorn Data

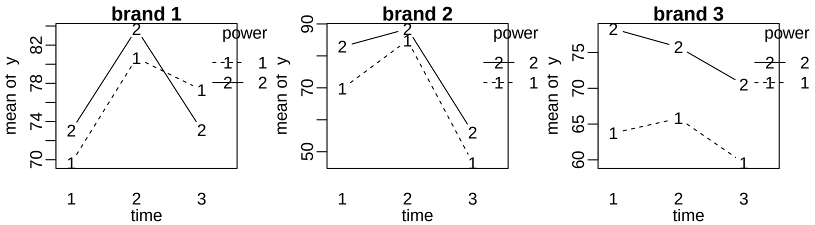
```r
with(subset(popcorn,brand==1),
   interaction.plot(time,power,y,type="b", main="brand 1"))
with(subset(popcorn,brand==2),
   interaction.plot(time,power,y,type="b", main="brand 2"))
with(subset(popcorn,brand==3),
   interaction.plot(time,power,y,type="b", main="brand 3"))
```



- Are there signs of `time:power` interactions?
- Does popping `time` has a greater effect for brand 1 or 2?

3-Way Interaction Plots — Popcorn Data

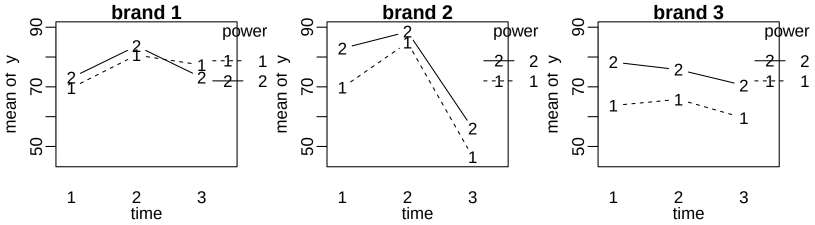
```
with(subset(popcorn,brand==1),
   interaction.plot(time,power,y,type="b", main="brand 1"))
with(subset(popcorn,brand==2),
   interaction.plot(time,power,y,type="b", main="brand 2"))
with(subset(popcorn,brand==3),
   interaction.plot(time,power,y,type="b", main="brand 3"))
```



- Are there signs of `time:power` interactions?
- Does popping `time` has a greater effect for brand 1 or 2?

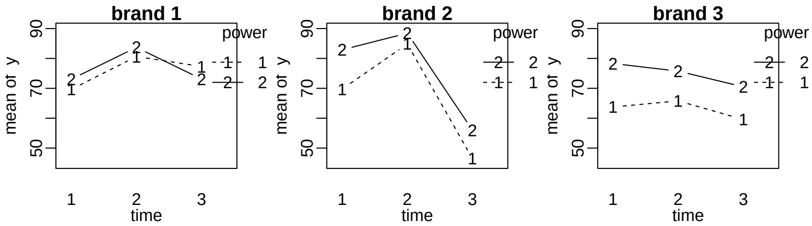
The 3 plots are on different y-axes.
Better put them on the same y-axis before comparison

```r
with(subset(popcorn,brand==1),
  interaction.plot(time,power,y,type="b",main="brand 1",ylim=c(45,90)))
with(subset(popcorn,brand==2),
  interaction.plot(time,power,y,type="b",main="brand 2",ylim=c(45,90)))
with(subset(popcorn,brand==3),
  interaction.plot(time,power,y,type="b",main="brand 3",ylim=c(45,90)))
```



- Greater time effect for Brand 2 (steeper lines) than Brands 1 & 3
  - higher popping % if popped 4.5 mins than 5 mins for Brand 2
  - signs of `brand:time` interactions
- If not placed on the same y-scale, Brand 1 seem to have greater `time` effect, which is not true

Popcorn Data: Are There 3-Way Interactions?



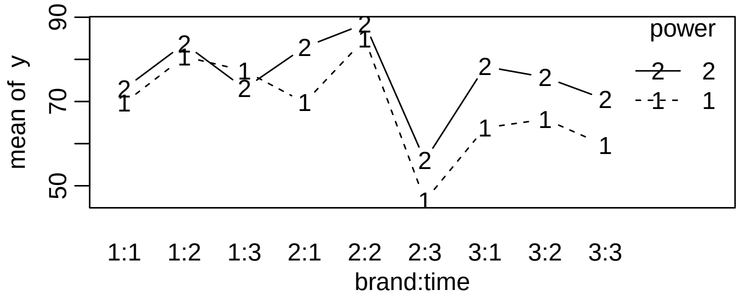
Placed on the same y-axis, the 2 lines appear closer to parallel, for all 3 brands.

- little `power:time` interactions for each brand
- Lines in an interaction plot may not be exactly parallel due to noise even if there is no interaction.

We can hence conclude there is little `brand:power:time` interactions as `power:time` interactions change little with `brand`

One can merge all three plots into one.

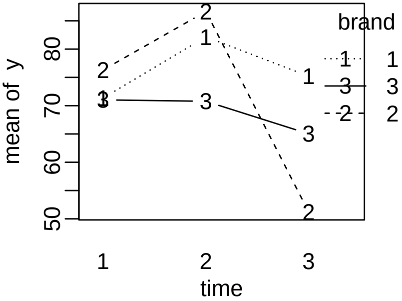
```
with(popcorn, interaction.plot(brand:time, power, y,
type="b"))
```



2-Way Interaction Plots — Popcorn Data

If one just check the two-way interaction plot between `time` and `brand`, the information of `power` would be ignored. The 3 lines below the lines the 3 brands averaged over the two levels of power.

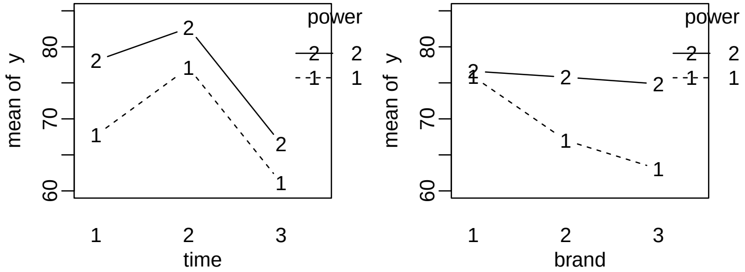
```
with(popcorn,interaction.plot(time, brand, y, type="b"))
```



Some evidence of `brand:time` interactions.

2-Way Interaction Plots — Popcorn Data

```
with(popcorn,interaction.plot(time, power, y, type="b", ylim=c(60,85)))
with(popcorn,interaction.plot(brand, power, y, type="b", ylim=c(60,85))
```



- Do the `time` main effects appear significant?
- How about the `power` main effect ?
- `brand` main effect ?
- `time:power` interaction?
- `brand:power` interaction?

Parameter Estimates — Popcorn Data

$\bar{y}_{ijk\bullet}$ (i = brand, j = power, k = time)

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

$\bar{y}_{ij\bullet\bullet}$:

```
mean(y ~ brand+power, data=popcorn)
  1.1   2.1   3.1   1.2   2.2   3.2
75.87 67.02 63.03 76.62 75.78 74.85
```

$\bar{y}_{i\bullet k\bullet}$

```
mean(y ~ brand+time, data=popcorn)
  1.1   2.1   3.1   1.2   2.2   3.2   1.3   2.3   3.3
71.35 76.33 71.03 82.17 86.65 70.78 75.20 51.23 65.03
```

$\bar{y}_{\bullet jk\bullet}$

```
mean(y ~ power+time, data=popcorn)
  1.1   2.1   1.2   2.2   1.3   2.3
67.72 78.08 77.10 82.63 61.10 66.53
```

$\bar{y}_{i\bullet\bullet\bullet}$:

```
mean(y ~ brand, data=popcorn)
    1     2     3
76.24 71.40 68.94
```

$\bar{y}_{\bullet j\bullet\bullet}$

```
mean(y ~ power, data=popcorn)
    1     2
68.64 75.75
```

$\bar{y}_{\bullet\bullet k\bullet}$

```
mean(y ~ time, data=popcorn)
    1     2     3
72.90 79.87 63.82
```

$\bar{y}_{\bullet\bullet\bullet\bullet}$ (grand mean)

```
mean(y ~ 1, data=popcorn)
    1
72.19
```

Parameter Estimates

For the full model with all 2-way and 3-way interactions

$$y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \beta\gamma_{jk} + \alpha\gamma_{ij} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell},$$

estimates for main effects under the zero-sum constraints are

$\widehat{\alpha}_i = \bar{y}_{i\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$:

```
mean(y ~ brand, data=popcorn)-mean(y ~ 1, data=popcorn)
      1       2       3
 4.0472 -0.7944 -3.2528
```

$\widehat{\beta}_j = \bar{y}_{\bullet j\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$

```
mean(y ~ power, data=popcorn)-mean(y ~ 1, data=popcorn)
     1      2
-3.556  3.556
```

$\widehat{\gamma}_k = \bar{y}_{\bullet\bullet k\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$

```
mean(y ~ time, data=popcorn)-mean(y ~ 1, data=popcorn)
      1       2       3
 0.7056  7.6722 -8.3778
```

```
mean(y ~ power+time, data=popcorn)
  1.1   2.1   1.2   2.2   1.3   2.3
67.72 78.08 77.10 82.63 61.10 66.53
mean(y ~ power, data=popcorn)
    1     2
68.64 75.75
mean(y ~ time, data=popcorn)
    1     2     3
72.90 79.87 63.82
```

$$\begin{aligned}\widehat{\beta\gamma}_{11} &= \bar{y}_{\bullet 11\bullet} - \bar{y}_{\bullet 1\bullet\bullet} - \bar{y}_{\bullet\bullet 1\bullet} + \bar{y}_{\bullet\bullet\bullet\bullet} \\ &\approx 67.72 - 68.64 - 72.9 + 72.19 = -1.63 \\ \widehat{\beta\gamma}_{12} &= \bar{y}_{\bullet 12\bullet} - \bar{y}_{\bullet 1\bullet\bullet} - \bar{y}_{\bullet\bullet 2\bullet} + \bar{y}_{\bullet\bullet\bullet\bullet} \\ &\approx 77.10 - 68.64 - 79.87 + 72.19 = 0.78\end{aligned}$$

Other $\widehat{\beta}\gamma_{ij}$'s can be obtained by the zero-sum constraints

$$\widehat{\beta\gamma}_{13} = -(\widehat{\beta\gamma}_{11} + \widehat{\beta\gamma}_{12}) \approx -1.63 + 0.78 = -0.85$$

$$\widehat{\beta\gamma}_{21} = -\widehat{\beta\gamma}_{11} \approx 1.63, \quad \widehat{\beta\gamma}_{22} = -\widehat{\beta\gamma}_{12} \approx -0.78, \quad \widehat{\beta\gamma}_{23} = -\widehat{\beta\gamma}_{13} \approx 0.85$$

$$
\begin{aligned}
\widehat{\alpha\beta\gamma}_{121} &= \bar{y}_{121\bullet} - \bar{y}_{12\bullet\bullet} - \bar{y}_{\bullet 21\bullet} - \bar{y}_{1\bullet 1\bullet} + \bar{y}_{1\bullet\bullet\bullet} + \bar{y}_{\bullet 2\bullet\bullet} + \bar{y}_{\bullet\bullet 1\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet} \\
&\approx 73.05 - 76.62 - 78.08 - 71.35 + 76.24 + 75.75 + 72.9 - 72.19 \\
&= -0.30
\end{aligned}
$$

Other parameters can be estimated similarly.

Double-checking our calculation of parameter estimates in R:

```
contrasts(popcorn$brand) = contr.sum(3)
contrasts(popcorn$power) = contr.sum(2)
contrasts(popcorn$time) = contr.sum(3)

# must re-fit model to update coefficients
lm1 = lm(y ~ brand*power*time, data=popcorn)
lm1$coef
        (Intercept)              brand1              brand2
            72.1944              4.0472             -0.7944
             power1               time1               time2
            -3.5556              0.7056              7.6722
      brand1:power1       brand2:power1        brand1:time1
             3.1806             -0.8278             -5.5972
       brand2:time1        brand1:time2        brand2:time2
             4.2194             -1.7389              7.5778
       power1:time1        power1:time2 brand1:power1:time1
            -1.6278              0.7889              0.3028
brand2:power1:time1 brand1:power1:time2 brand2:power1:time2
            -0.5639             -1.9389              1.7944
```

Model Formula in R

The R command for fitting the full 3-way model

$$y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \beta\gamma_{jk} + \alpha\gamma_{ij} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell}$$

is

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

A simpler syntax is

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

The term `brand*power*time` and `brand:power:time` both mean the 3-way interaction terms $\alpha\beta\gamma_{ijk}$, but

- `brand*power*time` will automatically include all relevant main effects and lower order interactions in the model.
- `brand:power:time` will not include the lower order terms

Sum of Squares — Popcorn Data

$$
\begin{aligned}
SS_B &= \sum\nolimits_{ijk\ell} (\widehat{\beta}_j)^2 = acn \sum\nolimits_j (\widehat{\beta}_j)^2 \\
&\approx (3)(3)(2)(3.556^2 + (-3.556)^2) \approx 455.11 \\
SS_C &= \sum\nolimits_{ijk\ell} (\widehat{\gamma}_k)^2 = abn \sum\nolimits_k (\widehat{\gamma}_k)^2 \\
&\approx (3)(2)(2)(0.7056^2 + 7.6722^2 + (-8.3778)^2) \approx 1554.58 \\
SS_{BC} &= \sum\nolimits_{ijk\ell} (\widehat{\beta\gamma}_{jk})^2 = an \sum\nolimits_{jk} (\widehat{\beta\gamma}_{jk})^2 \\
&\approx (3)(2)((-1.63)^2 + 0.78^2 + (-0.85)^2 + 1.63^2 + (-0.78)^2 + 0.85^2) \\
&\approx 47.85
\end{aligned}
$$

and so on

ANOVA Table — Popcorn Data

```
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
```

Only `power` and `time` main effects, and the `brand:time` interactions are significant.

Can I fit a model like $y_{ijk\ell} = \mu + \beta_j + \gamma_k + \alpha\gamma_{ik} + \varepsilon_{ijk\ell}$?

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

Hierarchy

Hierarchy

A model is *hierarchical* if any term in the model implies the presence of all the composite lower-order terms.

- $y_{ijk\ell} = \mu + \alpha_i + \beta_j + \beta\gamma_{jk} + \varepsilon_{ijk\ell}$ is not hierarchical because including the term $\beta\gamma_{jk}$ must includes both β_j and γ_k.
- $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ is hierarchical.
- A hierarchical model with a term $\alpha\beta\gamma_{ijk}$ must also include:
  - the relevant main effects: $\alpha_i + \beta_j + \gamma_k$
  - and the included two-way effects: $\alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk}$.

Hierarchy

Unless having a specific reason, we should stick to hierarchical models.

- This is because a k-way interaction in defined upon $(k-1)$-way interactions. It is strange to consider a ABC interaction while claiming A and B have no 2-way interaction.
- E.g., when we say there are no AB interactions, we also imply that there are no higher order interactions that involve AB interactions, like ABD interactions, or ABCD interactions.

Why Maintaining Hierarchy?

Let's consider a model for a 2×2 factorial design.

$$\begin{aligned} y_{ijk} &= \mu_{ij} + \varepsilon_{ijk} \\ &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \end{aligned}$$

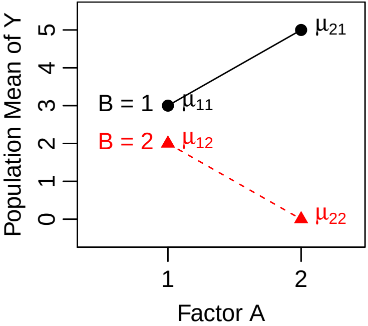
If $\alpha_1 = \alpha_2 = 0$, but $\alpha\beta_{11} \neq 0$, can Factor A have any effect on the response? Consider the example below.

	$B = 1$	$B = 2$	Mean
$A = 1$	$\mu_{11} = 3$	$\mu_{12} = 2$	$\mu_{1\bullet} = 2.5$
$A = 2$	$\mu_{21} = 5$	$\mu_{22} = 0$	$\mu_{2\bullet} = 2.5$
Mean	$\mu_{\bullet 1} = 4$	$\mu_{\bullet 2} = 1$	$\mu_{\bullet\bullet} = 2.5$

Under the zero-sum constraint,

$$\alpha_i = \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet} = 2.5 - 2.5 = 0$$

for $i = 1, 2$.
Clearly $\alpha\beta_{ij} \neq 0$ as the lines are not parallel.



Back to the Popcorn Data

Here is a hierarchical model that leaves out all insignificant terms.

```
lm3 = lm(y ~ brand + power + time + brand:time, data=popcorn)
anova(lm3)
Analysis of Variance Table

Response: y
           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
```

Cannot leave out the insignificant `brand` main effects since it involves in the significant the two-way interaction `brand:time`

The SS's and d.f.'s of the left-out terms are **pooled into the SSE and the df of error** while the SS's and d.f's of the remaining stay unchanged.

More On Model Formula in R (1)

Instead of writing terms explicitly in the model formula

```
lm3 = lm(y ~ brand + power + time + brand:time, data=popcorn)
```

Here is a simpler expression for the same model. R will automatically create the smallest hierarchical model that include `brand:time` interactions.

```
lm3a = lm(y ~ power + brand*time, data=popcorn)
anova(lm3a)
Analysis of Variance Table

Response: y
           Df   Sum Sq Mean Sq  F value     Pr(>F)
power       1  455.111 455.111  6.33129 0.01837031
brand       2  331.101 165.550  2.30306 0.11999055
time        2 1554.576 777.288 10.81326 0.00038249
brand:time  4 1433.858 358.464  4.98679 0.00405232
Residuals  26 1868.954  71.883
```

More On Model Formula in R (2)

To fit a model with all two-way interactions but no 3-way interaction, one can explicitly write down every term

```
lm(y ~ A + B + C + A:B + B:C + A:C)
```

Another way to obtain everything up to the 2-way interactions

```
lm(y ~ (A + B + C)^2)
```

Or one can "leave out" the 3-way interactions

```
lm(y ~ A*B*C - A:B:C, data=popcorn)
```