

STAT 222 Lecture 11-12
Chapter 6 Two-Way Factorial Designs

Yibi Huang

Example: Sprouting Barley (p.166 in Oehlert)

Brewer's malt is produced from germinating barley, so brewers like to know under what conditions they should germinate their barley. The following is part of an experiment on barley germination.

- ▶ 30 lots of barley seeds, 100 seeds per lot, are randomly divided into 10 groups of 3 lots
- ▶ Each group receives a treatment according to
 - ▶ water amount used in germination — 4 ml or 8 ml
 - ▶ age of seeds in weeks after harvest — 1, 3, 6, 9, or 12
- ▶ Response: # of seeds germinated

water	Age of Seeds (weeks)				
	1	3	6	9	12
4(ml)	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
8(ml)	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25

Basic Terminology

The sprouting barley experiment has 10 treatments. The 10 treatments has a *factorial structure*.

- ▶ A *factor* is an experimentally adjustable variable, e.g. water amount used in germination, age of seeds in weeks after harvest, ...
- ▶ Factors have *levels*, e.g.
water amount is a factor with 2 levels (4 ml or 8 ml)
age of seeds is a factor with 5 levels (1, 3, 6, 9, 12 weeks)
- ▶ A treatment is a *combination of factors*.
In the barley experiment, the treatments are the 2×5 combinations of the possible levels of the two factors

(4ml, 1 wk) (4ml, 3 wks) (4ml, 6 wks) (4ml, 9 wks) (4ml, 12 wks)
(8ml, 1 wk) (8ml, 3 wks) (8ml, 6 wks) (8ml, 9 wks) (8ml, 12 wks)

Full k -Way Factorial Design

- ▶ Consider k factors with respectively L_1, L_2, \dots, L_k levels, a *full k -way factorial design* include all the $L_1 \times L_2 \times \dots \times L_k$ combination of the k factors as treatments.
- ▶ A factorial design is said to be *balanced* if all the treatment groups have the same number of *replicates*. Otherwise, the design is *unbalanced*.
 - ▶ Question: How many units are there in a 3×2 design with 4 replicates?
- ▶ Balanced designs have many advantages, but not always necessary — sometimes if a unit fails (ex, a test tube gets dropped) we might end up with unbalanced results even if the original design was balanced

Data for a Two-Way $a \times b$ Design with n Replicates

	B -level 1	B -level 2		B -level b
A-level 1	y_{111}	y_{121}		y_{1b1}
	y_{112}	y_{122}	...	y_{1b2}
	\vdots	\vdots	...	\vdots
	y_{11n}	y_{12n}		y_{1bn}
A-level 2	y_{211}	y_{221}		y_{2b1}
	y_{212}	y_{222}	...	y_{2b2}
	\vdots	\vdots	...	\vdots
	y_{21n}	y_{22n}		y_{2bn}
\vdots	\vdots	\ddots	\vdots	
\vdots	\vdots		\vdots	
A-level a	y_{a11}	y_{a21}		y_{ab1}
	y_{a12}	y_{a22}	...	y_{ab2}
	\vdots	\vdots	...	\vdots
	y_{a1n}	y_{a2n}		y_{abn}

Means Model for a Two-Way Factorial Design

For a $a \times b$ two-way factorial experiment with n replicates

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

- ▶ y_{ijk} = the k th replicate in the treatment formed from the i th level of factor A and j th level of factor B
- ▶ ε_{ijk} 's are i.i.d. $N(0, \sigma^2)$
- ▶ μ_{ij} = the mean response in the treatment formed from the i th level of factor A and j th level of factor B
- ▶ The means model regards the 2-way factorial design as a completely randomized design (CRD) with $a \times b$ treatments, ignoring the factorial structure of the treatments.

Two-Way Interaction Contrast

The two-way interaction contrast between level (i_1, i_2) of factor A and level (j_1, j_2) of factor B is defined as

$$C = \mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2},$$

which has two interpretations.

$$\begin{aligned} C &= \mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} \\ &= \underbrace{(\mu_{i_1 j_1} - \mu_{i_1 j_2})}_{\text{effect of changing B from } j_1 \text{ to } j_2 \text{ when A is fixed at } i_1} - \underbrace{(\mu_{i_2 j_1} - \mu_{i_2 j_2})}_{\text{effect of changing B from } j_1 \text{ to } j_2 \text{ when A is fixed at } i_2} \\ &= \underbrace{(\mu_{i_1 j_1} - \mu_{i_2 j_1})}_{\text{effect of changing A from } i_1 \text{ to } i_2 \text{ when B is fixed at } j_1} - \underbrace{(\mu_{i_1 j_2} - \mu_{i_2 j_2})}_{\text{effect of changing A from } i_1 \text{ to } i_2 \text{ when B is fixed at } j_2} \end{aligned}$$

Two-Way Interaction

We say factor A and factor B have *no interaction* if and only if the two-way interaction contrast between any two levels of A and any two levels of B is 0, i.e.,

$$\mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} = 0 \quad \text{for all } i_1, i_2, j_1, j_2,$$

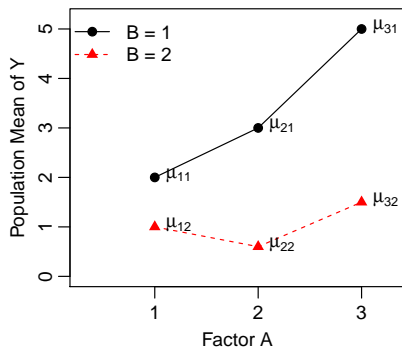
which has two interpretations:

- ▶ effect of A on Y doesn't change with the levels of B , and
- ▶ effect of B on Y doesn't change with the levels of A

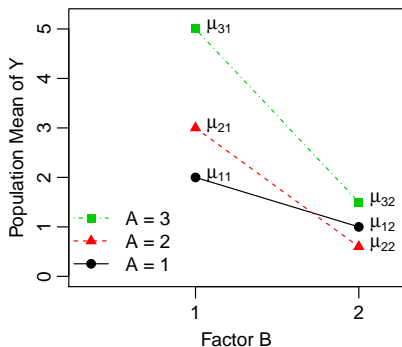
Conversely, two factors A and B are said to have *two-way interactions* if the effect of A on Y changes with the levels of B , or the effect of B on Y changes with the levels of A .

Interaction Plots

Plotting cell means (μ_{ij}) against levels of one factor (A or B), with different lines for the other factor (B or A)



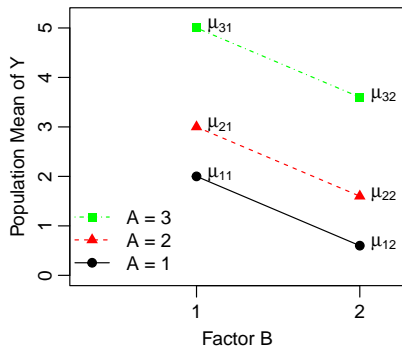
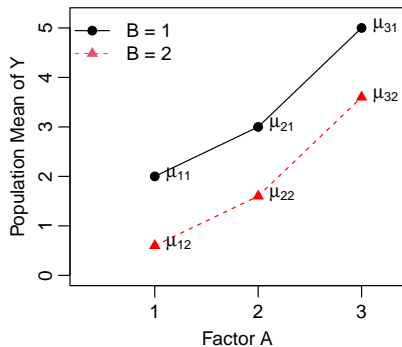
A is on the x-axis
B is the trace factor



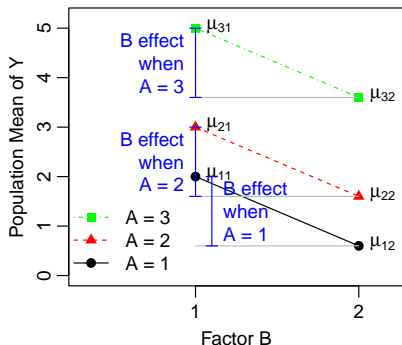
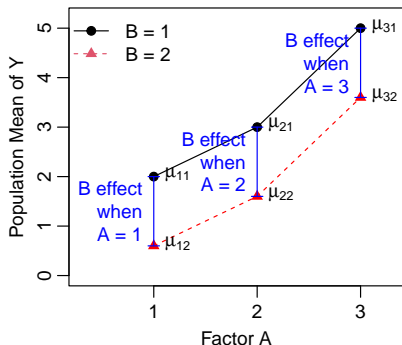
B is on the x-axis
A is the trace factor

The two interaction plots convey the same information.

Parallel Lines Indicate No Interaction (1)

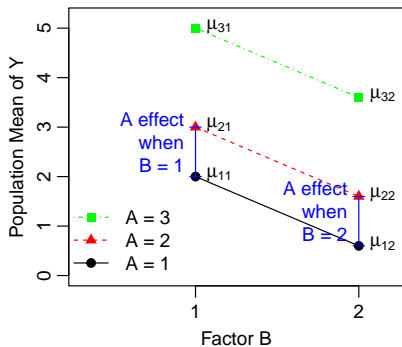
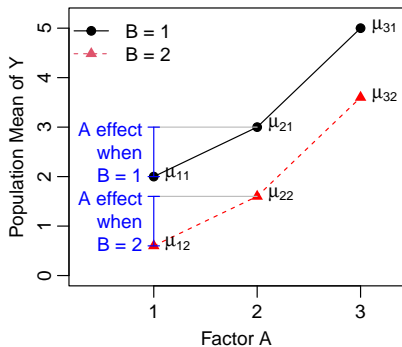


Parallel Lines Indicate No Interaction (2)



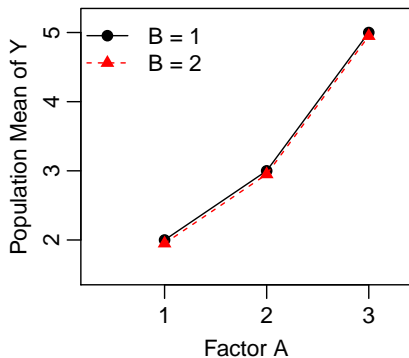
Effect of B on Y doesn't change with levels of A

Parallel Lines Indicate No Interaction (3)



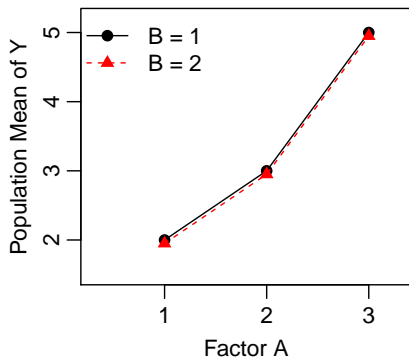
Effect of A on Y doesn't change with levels of B

What does the interaction plot below tell us? (1)



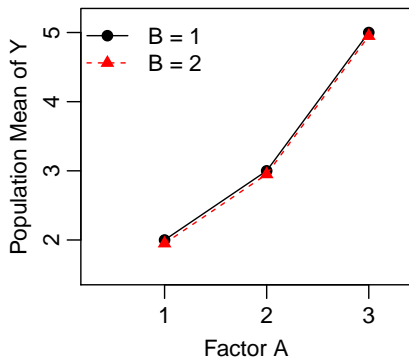
► No AB interaction

What does the interaction plot below tell us? (1)



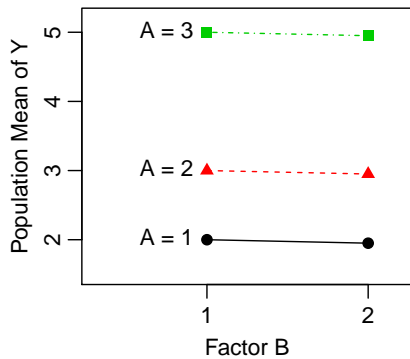
- ▶ No AB interaction
- ▶ B has no effect on Y since there is **no gap** between lines

What does the interaction plot below tell us? (1)

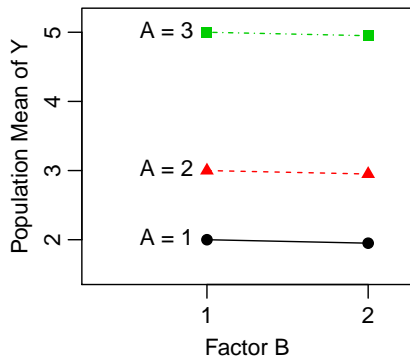


- ▶ No AB interaction
- ▶ B has no effect on Y since there is **no gap** between lines
- ▶ A has some effect on Y since the lines are **not horizontal**

What does the interaction plot below tell us? (2)

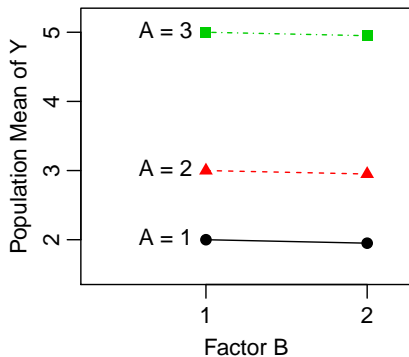


What does the interaction plot below tell us? (2)



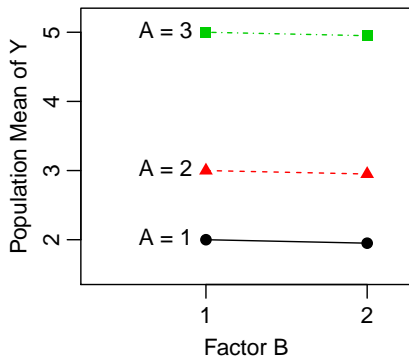
► No AB interaction

What does the interaction plot below tell us? (2)



- ▶ No AB interaction
- ▶ B has no effect on Y since the lines are horizontal

What does the interaction plot below tell us? (2)



- ▶ No AB interaction
- ▶ B has no effect on Y since the lines are horizontal
- ▶ A has some effect on Y since there are gaps between lines

Interaction Plots for the Sprouting Barley Study

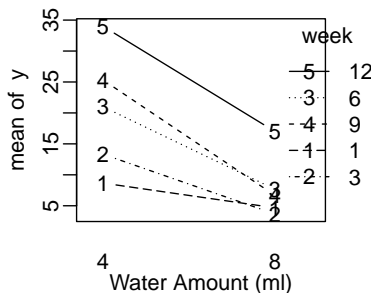
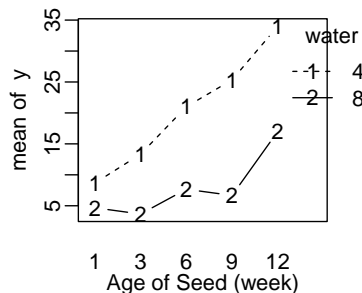
In reality, the population means μ_{ij} are not observable. Interaction plots are made using sample means $\bar{y}_{ij\bullet}$ rather than population means μ_{ij} .

y_{ijk}	Age of Seeds (weeks)				
	1	3	6	9	12
water 4(ml)	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
water 8(ml)	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25

sample means $\bar{y}_{ij\bullet}$	Age of Seeds (weeks)				
	1	3	6	9	12
water 4(ml)	8.67	13.33	21.00	25.33	34.00
water 8(ml)	4.67	3.67	7.67	6.67	17.00

Interaction Plots in R

```
barley = read.table(  
  "http://www.stat.uchicago.edu/~yibi/s222/SproutingBarley.txt",h=T)  
with(barley, interaction.plot(week,water,y,type="b",  
  xlab="Age of Seed (week)"))  
with(barley, interaction.plot(water,week,y,type="b",  
  xlab="Water Amount (ml)"))
```

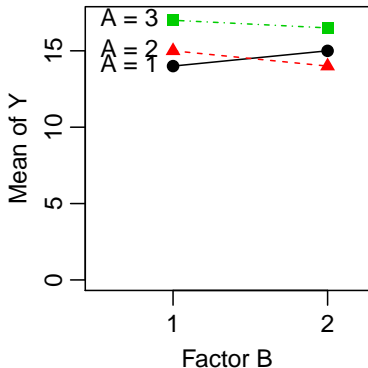
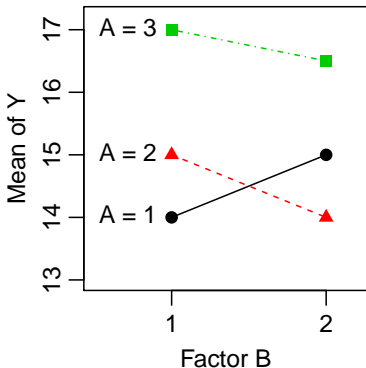


Lines in the interaction plots might not be exactly parallel even if the two factors have no interaction since $\bar{y}_{ij\bullet} \neq \mu_{ij}$.

The less parallel the lines, the stronger the evidence of interactions.

“Parallel” Or Not Is Affected by The Y-Scale

Please note that the y-scale might affect your perception of whether the lines are “parallel” or not.



Check the Y-scale and see if the change in the slopes is big enough to be important.

Additive Model

An **additive model** or **main-effect model** for two-way factorial data is as follows

$$y_{ijk} = \mu + \underbrace{\alpha_i}_{\text{A main effect}} + \underbrace{\beta_j}_{\text{B main effect}} + \varepsilon_{ijk} \quad \text{for } \begin{cases} i = 1, \dots, a, \\ j = 1, \dots, b, \\ k = 1, \dots, n. \end{cases}$$

- ▶ The additive model takes the factorial structure of the $a \times b$ treatments into account
- ▶ The additive model is nested in the means model

$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ since the means model will become the additive model if

$$\mu_{ij} = \mu + \alpha_i + \beta_j \quad \text{for all } i, j.$$

Additive Model Assumes No Interactions

If the additive model $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ is true, then

$$\mu_{ij} = \mu + \alpha_i + \beta_j \quad \text{for all } i, j,$$

we have

$$\begin{aligned} & \mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} \\ &= (\mu + \alpha_{i_1} + \beta_{j_1}) - (\mu + \alpha_{i_1} + \beta_{j_2}) \\ & \quad - (\mu + \alpha_{i_2} + \beta_{j_1}) + (\mu + \alpha_{i_2} + \beta_{j_2}) \\ &= 0 \end{aligned}$$

for all i_1, i_2, j_1, j_2 . Thus the two factors have no interaction.

Additive Model Assumes No Interactions

If the additive model $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ is true, then

$$\mu_{ij} = \mu + \alpha_i + \beta_j \quad \text{for all } i, j,$$

we have

$$\begin{aligned} & \mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} \\ &= (\mu + \alpha_{i_1} + \beta_{j_1}) - (\mu + \alpha_{i_1} + \beta_{j_2}) \\ & \quad - (\mu + \alpha_{i_2} + \beta_{j_1}) + (\mu + \alpha_{i_2} + \beta_{j_2}) \\ &= 0 \end{aligned}$$

for all i_1, i_2, j_1, j_2 . Thus the two factors have no interaction.

However, under the means model, the two factors might have interactions.

$$\mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} \text{ might not be } 0.$$

Main-Effect-Interaction Model for 2-Way Factorial Designs

The main-effect-interaction model is an extension of the additive model that allows interactions

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad \text{for } \begin{cases} i = 1, \dots, a, \\ j = 1, \dots, b, \\ k = 1, \dots, n. \end{cases}$$

► $\alpha\beta_{ij}$ is a parameter by itself. $\alpha\beta_{ij} \neq \alpha_i \times \beta_j$; $\alpha\beta_{ij} \neq \alpha \times \beta_{ij}$

μ_{11}	μ_{12}	\cdots	μ_{1b}
μ_{21}	μ_{22}	\cdots	μ_{2b}
\vdots	\vdots	\ddots	\vdots
μ_{a1}	μ_{a2}	\cdots	μ_{ab}

 $=$

μ			
-------	--	--	--

 $+$

α_1
α_2
\vdots
α_a

 $+$

β_1	β_2	\cdots	β_b
-----------	-----------	----------	-----------

 $+$

$\alpha\beta_{11}$	$\alpha\beta_{12}$	\cdots	$\alpha\beta_{1b}$
$\alpha\beta_{21}$	$\alpha\beta_{22}$	\cdots	$\alpha\beta_{2b}$
\vdots	\vdots	\ddots	\vdots
$\alpha\beta_{a1}$	$\alpha\beta_{a2}$	\cdots	$\alpha\beta_{ab}$

Main-Effect-Interaction Model Is Overparameterized

- ▶ The main-effect-interaction model $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ is equivalent to the means model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$. They have identical predicted values, residuals, and SSE.

- ▶ For a two-way $a \times b$ design, the means model has ab parameters; the main-effect-interaction model has $1 + a + b + ab$ parameters

μ_{11}	μ_{12}	\cdots	μ_{1b}
μ_{21}	μ_{22}	\cdots	μ_{2b}
\vdots	\vdots		\vdots
μ_{a1}	μ_{a2}	\cdots	μ_{ab}

- ▶ 1 parameter μ
- ▶ a parameters for A main effects: $\alpha_1, \alpha_2, \dots, \alpha_a$
- ▶ b parameters for B main effects: $\beta_1, \beta_2, \dots, \beta_b$

- ▶ ab parameters for AB interactions:

$\alpha\beta_{11}$	$\alpha\beta_{12}$	\cdots	$\alpha\beta_{1b}$
$\alpha\beta_{21}$	$\alpha\beta_{22}$	\cdots	$\alpha\beta_{2b}$
\vdots	\vdots	\ddots	\vdots
$\alpha\beta_{a1}$	$\alpha\beta_{a2}$	\cdots	$\alpha\beta_{ab}$

- ▶ Two equivalent models should have **identical numbers of parameters**. The main-effect-interaction model is overparameterized, meaning its parameters cannot be uniquely determined unless we set **constraints** on them.

Baseline Constraints (1)

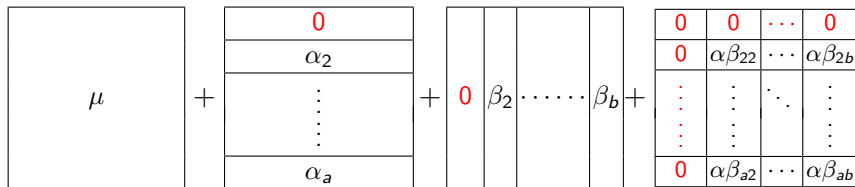
R by default uses the baseline constraints by setting all the parameters for the first level of a factor/interaction to 0,

$$\alpha_1 = 0, \quad \beta_1 = 0, \quad \alpha\beta_{1j} = \alpha\beta_{i1} = 0 \quad \text{for all } i, j.$$

Baseline Constraints (1)

R by default uses the baseline constraints by setting all the parameters for the first level of a factor/interaction to 0,

$$\alpha_1 = 0, \quad \beta_1 = 0, \quad \alpha\beta_{1j} = \alpha\beta_{i1} = 0 \quad \text{for all } i, j.$$



Effectively, there are

- ▶ 1 parameter μ ,
- ▶ $a - 1$ parameters for A main effects,
- ▶ $b - 1$ parameters for B main effects,
- ▶ $(a - 1)(b - 1)$ parameters for AB interactions.

In total, there are $1 + (a - 1) + (b - 1) + (a - 1)(b - 1) = ab$ parameters, same as the means model.

Baseline Constraints (2)

Under the baseline constraint: $\alpha_1 = 0$, $\beta_1 = 0$, $\alpha\beta_{1j} = \alpha\beta_{i1} = 0$

$$\blacktriangleright \mu_{11} = \mu + \underbrace{\alpha_1}_{=0} + \underbrace{\beta_1}_{=0} + \underbrace{\alpha\beta_{11}}_{=0} \Rightarrow \mu = \mu_{11}$$

$$\blacktriangleright \mu_{i1} = \mu + \alpha_i + \underbrace{\beta_1}_{=0} + \underbrace{\alpha\beta_{i1}}_{=0} \Rightarrow \alpha_i = \mu_{i1} - \mu = \mu_{i1} - \mu_{11}$$

▶ Thus α_i = effect of changing factor A from level 1 to level i on the mean of y , while fixing factor B at level 1

▶ Level 1 is the baseline level of factor A

$$\blacktriangleright \mu_{1j} = \mu + \underbrace{\alpha_1}_{=0} + \beta_j + \underbrace{\alpha\beta_{1j}}_{=0} \Rightarrow \beta_j = \mu_{1j} - \mu = \mu_{1j} - \mu_{11}$$

▶ Thus β_j = effect of changing factor B from level 1 to level j on the mean of y , while fixing factor A at level 1

▶ Level 1 is the baseline level of factor B

$$\begin{aligned}\blacktriangleright \alpha\beta_{ij} &= \mu_{ij} - \mu - \alpha_i - \beta_j \\ &= \mu_{ij} - \mu_{11} - (\mu_{i1} - \mu_{11}) - (\mu_{1j} - \mu_{11}) \\ &= \mu_{ij} - \mu_{i1} - \mu_{1j} + \mu_{11}\end{aligned}$$

Zero-Sum Constraints (1)

For factorial data, the more commonly used constraints are the zero-sum constraints:

$$\sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{i=1}^a \alpha\beta_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^b \alpha\beta_{ij} = 0 \text{ for all } i.$$

I.e., the row sums and column sums of the array $\{\alpha\beta_{ij}\}$ are all 0.

$\alpha\beta_{11}$	$\alpha\beta_{12}$	\cdots	$\alpha\beta_{1b}$	sum
$\alpha\beta_{21}$	$\alpha\beta_{22}$	\cdots	$\alpha\beta_{2b}$	0
\vdots	\vdots	\ddots	\vdots	\vdots
$\alpha\beta_{a1}$	$\alpha\beta_{a2}$	\cdots	$\alpha\beta_{ab}$	0
sum	0	0	\cdots	0

Zero-Sum Constraints (1)

For factorial data, the more commonly used constraints are the zero-sum constraints:

$$\sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{i=1}^a \alpha\beta_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^b \alpha\beta_{ij} = 0 \text{ for all } i.$$

I.e., the row sums and column sums of the array $\{\alpha\beta_{ij}\}$ are all 0.

There are hence effectively,

$\alpha\beta_{11}$	$\alpha\beta_{12}$	\cdots	$\alpha\beta_{1b}$	sum
$\alpha\beta_{21}$	$\alpha\beta_{22}$	\cdots	$\alpha\beta_{2b}$	0
\vdots	\vdots	\ddots	\vdots	\vdots
$\alpha\beta_{a1}$	$\alpha\beta_{a2}$	\cdots	$\alpha\beta_{ab}$	0
sum	0	0	\cdots	0

- ▶ 1 parameter μ ,
- ▶ $a - 1$ parameters for A main effects since $\alpha_a = -\sum_{i=1}^{a-1} \alpha_i$,
- ▶ $b - 1$ parameters for B main effects since $\beta_b = -\sum_{j=1}^{b-1} \beta_j$,
- ▶ $(a - 1)(b - 1)$ parameters for AB interactions since the last row and the last column of the $\{\alpha\beta_{ij}\}$ array can be determined from the zero-sum constraint.

In total, there are $1 + (a - 1) + (b - 1) + (a - 1)(b - 1) = ab$ parameters, same as the means model.

Zero-Sum Constraints (2)

Since $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, summing them over i , we get

$$\sum_{i=1}^a \mu_{ij} = a\mu + \underbrace{\sum_{i=1}^a \alpha_i}_{=0} + a\beta_j + \underbrace{\sum_{i=1}^a \alpha\beta_{ij}}_{=0}, \Rightarrow \mu + \beta_j = \frac{1}{a} \sum_{i=1}^a \mu_{ij} = \bar{\mu}_{\bullet j}$$

Zero-Sum Constraints (2)

Since $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, summing them over i , we get

$$\sum_{i=1}^a \mu_{ij} = a\mu + \underbrace{\sum_{i=1}^a \alpha_i}_{=0} + a\beta_j + \underbrace{\sum_{i=1}^a \alpha\beta_{ij}}_{=0}, \Rightarrow \mu + \beta_j = \frac{1}{a} \sum_{i=1}^a \mu_{ij} = \bar{\mu}_{\bullet j}$$

Likewise, summing $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$ over j , we get

$$\sum_{j=1}^b \mu_{ij} = b\mu + b\alpha_i + \underbrace{\sum_{j=1}^b \beta_j}_{=0} + \underbrace{\sum_{j=1}^b \alpha\beta_{ij}}_{=0}, \Rightarrow \mu + \alpha_i = \frac{1}{b} \sum_{j=1}^b \mu_{ij} = \bar{\mu}_{i \bullet}$$

Zero-Sum Constraints (2)

Since $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, summing them over i , we get

$$\sum_{i=1}^a \mu_{ij} = a\mu + \underbrace{\sum_{i=1}^a \alpha_i}_{=0} + a\beta_j + \underbrace{\sum_{i=1}^a \alpha\beta_{ij}}_{=0}, \Rightarrow \mu + \beta_j = \frac{1}{a} \sum_{i=1}^a \mu_{ij} = \bar{\mu}_{\bullet j}$$

Likewise, summing $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$ over j , we get

$$\sum_{j=1}^b \mu_{ij} = b\mu + b\alpha_i + \underbrace{\sum_{j=1}^b \beta_j}_{=0} + \underbrace{\sum_{j=1}^b \alpha\beta_{ij}}_{=0}, \Rightarrow \mu + \alpha_i = \frac{1}{b} \sum_{j=1}^b \mu_{ij} = \bar{\mu}_{i\bullet}$$

Summing $\mu + \alpha_i = \bar{\mu}_{i\bullet}$ over i , we get

$$a\mu + \underbrace{\sum_{i=1}^a \alpha_i}_{=0} = \sum_{i=1}^a \bar{\mu}_{i\bullet} = \sum_{i=1}^a \frac{1}{b} \sum_{j=1}^b \mu_{ij} \Rightarrow \mu = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \mu_{ij} = \bar{\mu}_{\bullet\bullet}$$

Zero-Sum Constraints (3)

Under the zero-sum constraint, parameters in the means model and the main-effect-interaction model are related as follows

$$\mu = \bar{\mu}_{\bullet\bullet} = \text{overall mean}$$

$$\alpha_i = \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet} = \text{row mean} - \text{overall mean}$$

$$\beta_j = \bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet} = \text{column mean} - \text{overall mean}$$

$$\alpha\beta_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$$

$$= \mu_{ij} - \bar{\mu}_{\bullet\bullet} - (\bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet}) - (\bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet})$$

$$= \mu_{ij} - \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet}$$

$$= \text{cell mean} - \text{row mean} - \text{column mean} + \text{overall mean}$$

				row mean	
	μ_{11}	μ_{12}	\cdots	μ_{1b}	$\bar{\mu}_{1\bullet}$
	μ_{21}	μ_{22}	\cdots	μ_{2b}	$\bar{\mu}_{2\bullet}$
	\vdots	\vdots		\vdots	\vdots
	μ_{a1}	μ_{a2}	\cdots	μ_{ab}	$\bar{\mu}_{a\bullet}$
column mean	$\bar{\mu}_{\bullet 1}$	$\bar{\mu}_{\bullet 2}$	\cdots	$\bar{\mu}_{\bullet b}$	$\bar{\mu}_{\bullet\bullet} = \text{overall mean}$

Estimation of Parameters under Zero-Sum Constraints(1)

Parameter estimation in a *balanced* factorial design under the Zero-Sum constraint is straightforward. For

- ▶ $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ (means model)
- ▶ $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ (main-effect-interaction model)

the parameter estimates are

$$\hat{\mu}_{ij} = \bar{y}_{ij\bullet}$$

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet},$$

$$\hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet},$$

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

$$\hat{\alpha}\hat{\beta}_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}$$

Observe the estimates satisfy the zero-sum constraints:

$$\sum_{i=1}^a \hat{\alpha}_i = \sum_{j=1}^b \hat{\beta}_j = \sum_{i=1}^a \hat{\alpha}\hat{\beta}_{ij} = \sum_{j=1}^b \hat{\alpha}\hat{\beta}_{ij} = 0, \quad \text{for all } i, j.$$

Estimation of Parameters under Zero-Sum Constraints(2)

Since the design is balanced, for any of the reduced models below,

- ▶ $y_{ijk} = \mu + \varepsilon_{ijk} \dots \dots \dots$ (no main effects, no interaction)
- ▶ $y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk} \dots \dots \dots$ (main effects of A only)
- ▶ $y_{ijk} = \mu + \beta_j + \varepsilon_{ijk} \dots \dots \dots$ (main effects of B only)
- ▶ $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \dots \dots \dots$ (additive model)

the estimates of μ , α_i 's, and β_j 's under the zero-sum constraints are **identical** with those for the main-effects-interaction model:

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}, \quad \hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \quad \hat{\beta}_j = \bar{y}_{\bullet j \bullet} - \bar{y}_{\bullet\bullet\bullet}$$

If NOT balanced, the estimates will change with the model.

Recall in a regression model, the estimate of a coefficient will change with the presence of other covariates in the model.

Parameter Estimates Under the Baseline Constraints (May Skip)

Under the baseline constraints,

$$\alpha_1 = 0, \quad \beta_1 = 0, \quad \alpha\beta_{1j} = \alpha\beta_{i1} = 0 \quad \text{for all } i, j.$$

the least-square estimates for parameters in the 5 models are different (see below), even if the data is balanced.

Model Formula	Parameter Estimates			
	$\hat{\mu}$	$\hat{\alpha}_i$	$\hat{\beta}_j$	$\hat{\alpha}\hat{\beta}_{ij}$
$\mu + \alpha_i + \beta_j + \alpha\beta_{ij}$	$\bar{y}_{11\bullet}$	$\bar{y}_{i1\bullet} - \bar{y}_{11\bullet}$	$\bar{y}_{1j\bullet} - \bar{y}_{11\bullet}$	$\bar{y}_{ij\bullet} - \bar{y}_{i1\bullet} - \bar{y}_{1j\bullet} + \bar{y}_{11\bullet}$
$\mu + \alpha_i + \beta_j$	$\bar{y}_{1\bullet\bullet} + \bar{y}_{\bullet 1\bullet} - \bar{y}_{\bullet\bullet\bullet}$	$\bar{y}_{i\bullet\bullet} - \bar{y}_{1\bullet\bullet}$	$\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet 1\bullet}$	—
$\mu + \alpha_i$	$\bar{y}_{1\bullet\bullet}$	$\bar{y}_{i\bullet\bullet} - \bar{y}_{1\bullet\bullet}$	—	—
$\mu + \beta_j$	$\bar{y}_{\bullet 1\bullet}$	—	$\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet 1\bullet}$	—
μ	$\bar{y}_{\bullet\bullet\bullet}$	—	—	—

- ▶ Simplicity in the formulas for parameter estimates is the primary reason we adopt the zero-sum constraints for factorial data, even though R uses the baseline constraints
- ▶ **Don't memorize the formulas for the baseline constraints!**
- ▶ Models are not affected by the constraints imposed. The fitted values, residuals, df, SSE are not affected.

Fitted Values for a Main-Effect-Interaction Model

For a main-effect-interaction model, the fitted value for y_{ijk} under the zero-sum constraints is

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij} \\ &= \bar{y}_{\bullet\bullet\bullet} + (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) + (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}) \\ &\quad + (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}) \\ &= \bar{y}_{ij\bullet} = \text{cell mean}\end{aligned}$$

which is equal to the fitted value under the baseline constraints:

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij} \\ &= \bar{y}_{11\bullet} + (\bar{y}_{i1\bullet} - \bar{y}_{11\bullet}) + (\bar{y}_{1j\bullet} - \bar{y}_{11\bullet}) \\ &\quad + (\bar{y}_{ij\bullet} - \bar{y}_{i1\bullet} - \bar{y}_{1j\bullet} + \bar{y}_{11\bullet}) \\ &= \bar{y}_{ij\bullet} = \text{cell mean.}\end{aligned}$$

Fitted Values for an Additive Model

For an additive model (**no interaction**), the fitted value for y_{ijk} under the zero-sum constraints is

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \\ &= \bar{y}_{\bullet\bullet\bullet} + (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) + (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}) \\ &= \bar{y}_{i\bullet\bullet} + \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet} \\ &= \text{row mean} + \text{column mean} - \text{overall mean}\end{aligned}$$

which is equal to the fitted value under the baseline constraints:

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \\ &= (\bar{y}_{1\bullet\bullet} + \bar{y}_{\bullet 1\bullet} - \bar{y}_{\bullet\bullet\bullet}) + (\bar{y}_{i\bullet\bullet} - \bar{y}_{1\bullet\bullet}) + (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet 1\bullet}) \\ &= \bar{y}_{i\bullet\bullet} + \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet} \\ &= \text{row mean} + \text{column mean} - \text{overall mean}\end{aligned}$$

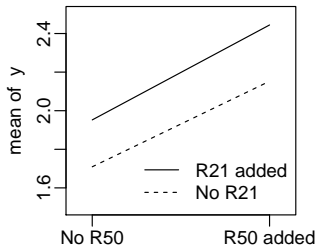
Bacteria in Cheese (p.178 in Oehlert)

- ▶ Factor A: Bacteria R50#10, added or not
- ▶ Factor B: Bacteria R21#2, added or not
- ▶ 3 replicates
- ▶ Response: total free amino acids in cheddar cheese after 56 days of ripening.

	No R21	R21 added
No R50	1.697 1.601 1.830	2.211 1.673 1.973
R50 added	2.032 2.017 2.409	2.091 2.255 2.987

⇒

	No R21	R21 added
No R50	$\bar{y}_{11\bullet} = 1.709$	$\bar{y}_{12\bullet} = 1.952$
R50 added	$\bar{y}_{21\bullet} = 2.153$	$\bar{y}_{22\bullet} = 2.444$



Is there interaction?

Bacteria in Cheese (p.178 in Oehlert)

	B-level 1	B-level 2	row mean
A-level 1	$\bar{y}_{11\bullet} = 1.709$	$\bar{y}_{12\bullet} = 1.952$	$\bar{y}_{1\bullet\bullet} = 1.831$
A-level 2	$\bar{y}_{21\bullet} = 2.153$	$\bar{y}_{22\bullet} = 2.444$	$\bar{y}_{2\bullet\bullet} = 2.299$
column mean	$\bar{y}_{\bullet 1\bullet} = 1.931$	$\bar{y}_{\bullet 2\bullet} = 2.198$	$\bar{y}_{\bullet\bullet\bullet} = 2.065$

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet} = 2.065$$

$$\hat{\alpha}_1 = \bar{y}_{1\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet} = 1.831 - 2.065 = -0.234$$

$$\hat{\beta}_1 = \bar{y}_{\bullet 1\bullet} - \bar{y}_{\bullet\bullet\bullet} = 1.931 - 2.065 = -0.134$$

$$\begin{aligned}\widehat{\alpha\beta}_{11} &= \bar{y}_{11\bullet} - \bar{y}_{1\bullet\bullet} - \bar{y}_{\bullet 1\bullet} + \bar{y}_{\bullet\bullet\bullet} \\ &= 1.709 - 1.831 - 1.931 + 2.065 = 0.012\end{aligned}$$

The estimates of all other parameters can be computed by the zero-sum constraints.

$$\hat{\alpha}_1 + \hat{\alpha}_2 = 0 \Rightarrow \hat{\alpha}_2 = -\hat{\alpha}_1 = 0.234$$

$$\hat{\beta}_1 + \hat{\beta}_2 = 0 \Rightarrow \hat{\beta}_2 = -\hat{\beta}_1 = 0.134$$

$$\widehat{\alpha\beta}_{11} + \widehat{\alpha\beta}_{12} = 0 \Rightarrow \widehat{\alpha\beta}_{12} = -\widehat{\alpha\beta}_{11} = -0.012$$

$$\widehat{\alpha\beta}_{11} + \widehat{\alpha\beta}_{21} = 0 \Rightarrow \widehat{\alpha\beta}_{21} = -\widehat{\alpha\beta}_{11} = -0.012$$

$$\widehat{\alpha\beta}_{12} + \widehat{\alpha\beta}_{22} = 0 \Rightarrow \widehat{\alpha\beta}_{22} = -\widehat{\alpha\beta}_{12} = 0.012$$

Finding Parameter Estimates in R

Note that R finding parameter estimates using the **baseline constraints** by default.

```
cheese = read.table(  
  "http://users.stat.umn.edu/~gary/book/fcdae.data/exmpl8.6", h=T)  
cheese$r50 = as.factor(cheese$r50)  
cheese$r21 = as.factor(cheese$r21)  
lmcheese = lm(y ~ r50 + r21 + r50*r21, data=cheese)  
lmcheese$coef  
(Intercept)          r502          r212    r502:r212  
    1.70933      0.44333      0.24300      0.04867
```

$$\hat{\mu} = \bar{y}_{11\bullet} \approx 1.709$$

$$\hat{\alpha}_2 = \bar{y}_{21\bullet} - \bar{y}_{11\bullet} \approx 2.153 - 1.709 = 0.444$$

$$\hat{\beta}_2 = \bar{y}_{12\bullet} - \bar{y}_{11\bullet} \approx 1.952 - 1.709 = 0.243$$

$$\begin{aligned}\hat{\alpha}_{22} &= \bar{y}_{22\bullet} - \bar{y}_{21\bullet} - \bar{y}_{12\bullet} + \bar{y}_{11\bullet} \\ &\approx 2.444 - 2.153 - 1.952 + 1.709 = 0.048\end{aligned}$$

$$\text{and } \hat{\alpha}_1 = \hat{\beta}_1 = \hat{\alpha}\hat{\beta}_{11} = \hat{\alpha}\hat{\beta}_{12} = \hat{\alpha}\hat{\beta}_{21} = 0.$$

	R21	
R50	No (B=1)	added (B=2)
No (A=1)	$\bar{y}_{11\bullet} = 1.709$	$\bar{y}_{12\bullet} = 1.952$
added (A=2)	$\bar{y}_{21\bullet} = 2.153$	$\bar{y}_{22\bullet} = 2.444$

How to Force R Using the Zero-Sum Constraints?

To force R using the zero-sum constraints, one needs to set the following

```
contrasts(cheese$r50) = contr.sum(2)
contrasts(cheese$r21) = contr.sum(2)
```

where the number **2** inside `contr.sum(2)` is the number of levels for the factor.

Next, one *must* fit the `lm()` model *again* to update the estimated coefficients.

```
lmcheese = lm(y ~ r50 + r21 + r50*r21, data=cheese)
lmcheese$coef
(Intercept)      r501      r211  r501:r211
  2.06467    -0.23383    -0.13367     0.01217
```

We get $\hat{\mu} \approx 2.065$, $\hat{\alpha}_1 \approx -0.234$, $\hat{\beta}_1 \approx -0.134$, $\hat{\alpha}\hat{\beta}_{11} \approx 0.012$ which match our calculations. Estimates for other parameters can be determined by the zero-sum constraints.

Sum of Squares for Balanced 2-Way Factorial Designs (1)

An balanced $a \times b$ two-way factorial design with n replicates is also a CRD with ab treatments, so the sum of squares identity remains valid.

$$SST = SS_{trt} + SSE$$

where

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2 \quad \text{and}$$

$$SS_{trt} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet})^2, \quad SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij\bullet})^2$$

d.f. for SST = total # of observations $- 1 = abn - 1$

d.f. for SS_{trt} = # of treatments $- 1 = ab - 1$

d.f. for SSE = total # of observations $-$ # of treatments
 $= abn - ab = ab(n - 1)$

Sum of Squares for Balanced 2-Way Factorial Designs (2)

As the ab treatments have a factorial structure, SS_{trt} can be decomposed further as

$$SS_{trt} = SS_A + SS_B + SS_{AB}$$

in which

SS	formula	d.f.
SS_A	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$
SS_B	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$	$b - 1$
SS_{AB}	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$(a-1)(b-1)$
SS_{trt}	$n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{...})^2$	$ab - 1$

Observe all the d.f.s for the SS of the main effects or interactions equal (number of parameters) – (number of constraint(s))

Sum of Squares for Balanced 2-Way Factorial Designs (3)

In summary

$$SST = SS_A + SS_B + SS_{AB} + SSE$$

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{\dots})^2$$

$$SS_A = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \underbrace{(\bar{y}_{i\bullet\bullet} - \bar{y}_{\dots})}_{\hat{\alpha}_i}^2 = bn \sum_{i=1}^a \hat{\alpha}_i^2$$

$$SS_B = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \underbrace{(\bar{y}_{\bullet j\bullet} - \bar{y}_{\dots})}_{\hat{\beta}_j}^2 = an \sum_{j=1}^b \hat{\beta}_j^2$$

$$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \underbrace{(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\dots})}_{\hat{\alpha}\hat{\beta}_{ij}}^2 = n \sum_{i=1}^a \sum_{j=1}^b \hat{\alpha}\hat{\beta}_{ij}^2$$

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij\bullet})^2$$

ANOVA Table for Balanced Two-Way Factorial Designs

Source	d.f.	SS	MS	F
Factor A	$a - 1$	SS_A	$MS_A = \frac{SS_A}{a-1}$	$F_A = \frac{MS_A}{MSE}$
Factor B	$b - 1$	SS_B	$MS_B = \frac{SS_B}{b-1}$	$F_B = \frac{MS_B}{MSE}$
AB Interaction	$(a-1)(b-1)$	SS_{AB}	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_{AB} = \frac{MS_{AB}}{MSE}$
Error	$ab(n-1)$	SSE	$MSE = \frac{SSE}{ab(n-1)}$	
Total	$abn - 1$	SST		

Questions of Interest in a 2-Way Factorial Design

1. Does factor A has an effect on the response?

E.g. does the age of seeds has an effect on germination?

$$\begin{cases} H_0 : \alpha_1 = \dots = \alpha_a = 0 \\ H_a : \text{not all } \alpha_i\text{'s} = 0, \end{cases} \Rightarrow F_A = \frac{MS_A}{MSE} \sim F_{a-1, ab(n-1)} \text{ under } H_0.$$

2. Does factor B has an effect on the response?

E.g. does the water amount has an effect on germination?

$$\begin{cases} H_0 : \beta_1 = \dots = \beta_b = 0 \\ H_a : \text{not all } \beta_i\text{'s} = 0, \end{cases} \Rightarrow F_B = \frac{MS_B}{MSE} \sim F_{b-1, ab(n-1)} \text{ under } H_0.$$

3. Does the effect of factor A interact with that of factor B?

E.g., does the effect of age change with water amount?

$$\begin{cases} H_0 : \alpha\beta_{ij} = 0 \text{ for all } i, j \\ H_a : \alpha\beta_{ij} \neq 0 \text{ for some } i, j \end{cases} \Rightarrow F_{AB} = \frac{MS_{AB}}{MSE} \sim F_{(a-1)(b-1), ab(n-1)} \text{ under } H_0.$$

Remark on the Test of Main Effects

Under the zero-sum constraints, $\alpha_j = \bar{\mu}_{j\bullet} - \bar{\mu}_{\bullet\bullet}$, the null hypothesis $H_0: \alpha_1 = \dots = \alpha_a = 0$ of the F -test of A main effects actually means

$$\bar{\mu}_{1\bullet} = \bar{\mu}_{2\bullet} = \dots = \bar{\mu}_{a\bullet}$$

which just means

*levels of A (averaged over the levels of B)
have the same average effect on the response.*

not necessarily means levels of A has the same effect on the response at each levels of B , i

$$\mu_{1j} = \mu_{2j} = \dots = \mu_{aj} \quad \text{for all } j = 1, 2, \dots, b.$$

unless A and B have no interactions.

Ditto for the F -test of B main effects.

A Model w/ 0 Main effects but Non-zero Interactions

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.

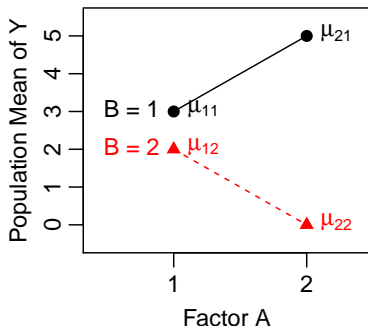
	B = 1	B = 2	Mean
A = 1	$\mu_{11} = 3$	$\mu_{12} = 2$	$\mu_{1\bullet} = 2.5$
A = 2	$\mu_{21} = 5$	$\mu_{22} = 0$	$\mu_{2\bullet} = 2.5$
Mean	$\mu_{\bullet 1} = 4$	$\mu_{\bullet 2} = 1$	$\mu_{\bullet\bullet} = 2.5$

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.



Example: Bacteria in Cheese (p.178 in Oehlert)

$$SS_A = bn \sum_{i=1}^a \hat{\alpha}_i^2 = 2 \times 3 \times [(-0.234)^2 + 0.234^2] = 0.656$$

$$SS_B = an \sum_{j=1}^b \hat{\beta}_j^2 = 2 \times 3 \times [(-0.134)^2 + 0.134^2] = 0.214$$

$$SS_{AB} = n \sum_{i=1}^a \sum_{j=1}^b \hat{\alpha}\hat{\beta}_{ij}^2 = 3 \times [0.012^2 \times 4] \approx 0.0017$$

Computing SSE needs more work. It is easier to compute the SST:

$$\begin{aligned} SST &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{\dots})^2 \\ &= (1.697 - 2.065)^2 + (1.601 - 2.065)^2 + (1.830 - 2.065)^2 \\ &\quad + \dots + (2.987 - 2.065)^2 = 1.598 \end{aligned}$$

Then we can get

$$\begin{aligned} SSE &= SST - SS_A - SS_B - SS_{AB} \\ &= 1.598 - 0.656 - 0.214 - 0.0018 = 0.726. \end{aligned}$$

Example: Bacteria in Cheese — ANOVA table

Source	d.f.	SS	MS	F-value	P-value
A(R50)	1	0.656	0.656	7.23	0.028
B(R21)	1	0.214	0.214	2.36	0.16
AB interaction	1	0.0017	0.0017	0.019	0.89
Error	8	0.726	0.091		
Total	11	1.598			

Only main effect A (Bacteria R50) is moderately significant.
Main effect B and interaction are not.

One can also obtain the ANOVA table in R.

```
lmcheese = lm(y ~ r50 + r21 + r50*r21, data=cheese)
anova(lmcheese)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
r50	1	0.65614	0.65614	7.2335	0.02752
r21	1	0.21440	0.21440	2.3636	0.16275
r50:r21	1	0.00178	0.00178	0.0196	0.89217
Residuals	8	0.72566	0.09071		

Recall Sprouting Barley Data

y_{ijk}	Age of Seeds (weeks)				
	1	3	6	9	12
water 4(ml)	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
water 8(ml)	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25

Cell means $\bar{y}_{ij\bullet}$	Age of Seeds (weeks)					Row means
	1	3	6	9	12	$\bar{y}_{i\bullet\bullet}$
water 4(ml)	8.67	13.33	21.00	25.33	34.00	20.47
water 8(ml)	4.67	3.67	7.67	6.67	17.00	7.93
Column means $\bar{y}_{\bullet j\bullet}$	6.67	8.50	14.33	16.00	25.50	$\bar{y}_{\bullet\bullet\bullet} = 14.2$ overall mean

Does water have an effect on germination?

Does the age of seeds have an effect?

Finding Row Means, Column Means, Cell Means in R

Overall mean $\bar{y}_{\dots} = \hat{\mu}$:

```
library(mosaic)
mean(~y, data=barley)
[1] 14.2
```

Row means $\bar{y}_{i\bullet\bullet}$ and Column means $\bar{y}_{\bullet j\bullet}$:

```
mean(y ~ water, data=barley)
      4      8
20.4667  7.9333
mean(y ~ week, data=barley)
      1      3      6      9      12
6.6667  8.5000 14.3333 16.0000 25.5000
```

Cell means ($\bar{y}_{ij\bullet}$, average of the 3 values in each cell):

```
mean(y ~ week+water, data=barley)
      1.4      3.4      6.4      9.4      12.4      1.8      3.8      6.8
8.6667 13.3333 21.0000 25.3333 34.0000  4.6667  3.6667  7.6667
      9.8      12.8
6.6667 17.0000
```


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

```
mean(y ~ water, data=barley)-mean(~y, data=barley)
      4      8
6.2667 -6.2667
```

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

```
mean(y ~ week, data=barley)-mean(~y, data=barley)
      1      3      6      9     12
-7.53333 -5.70000  0.13333  1.80000 11.30000
```

$$\widehat{\alpha\beta}_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}$$

```

cell.mean = matrix(mean(y ~ week+water, data=barley),nrow=2,byrow=T)
cell.mean
      [,1] [,2] [,3] [,4] [,5]
[1,] 8.6667 13.3333 21.0000 25.3333 34
[2,] 4.6667 3.6667 7.6667 6.6667 17
row.mean = mean(y ~ water, data=barley)%o%rep(1,5); row.mean
      [,1] [,2] [,3] [,4] [,5]
4 20.4667 20.4667 20.4667 20.4667 20.4667
8 7.9333 7.9333 7.9333 7.9333 7.9333
column.mean = rep(1,2)%o%mean(y ~ week, data=barley);column.mean
      1 3 6 9 12
[1,] 6.6667 8.5 14.333 16 25.5
[2,] 6.6667 8.5 14.333 16 25.5
overall.mean = mean(~y, data=barley)
cell.mean - row.mean - column.mean + overall.mean
      [,1] [,2] [,3] [,4] [,5]
4 -4.2667 -1.4333 0.4 3.0667 2.2333
8 4.2667 1.4333 -0.4 -3.0667 -2.2333

```

```

barley$weekfac = as.factor(barley$week)
barley$waterfac = as.factor(barley$water)
contrasts(barley$weekfac) = contr.sum(5)
contrasts(barley$waterfac) = contr.sum(2)
lmbarley = lm(y ~ waterfac + weekfac + waterfac*weekfac, data=barley)
coef(lmbarley)

```

(Intercept)	waterfac1	weekfac1
14.20000	6.26667	-7.53333
weekfac2	weekfac3	weekfac4
-5.70000	0.13333	1.80000
waterfac1:weekfac1	waterfac1:weekfac2	waterfac1:weekfac3
-4.26667	-1.43333	0.40000
waterfac1:weekfac4		
3.06667		

Observe that we get

$$\begin{aligned}
 \hat{\mu} &= 14.2, & \hat{\alpha}_1 &\approx 6.267, \\
 \hat{\beta}_1 &\approx -7.533, & \hat{\beta}_2 &= -5.7, & \hat{\beta}_3 &\approx 0.13, & \hat{\beta}_4 &= 1.8, \\
 \hat{\alpha}\hat{\beta}_{11} &\approx -4.266, & \hat{\alpha}\hat{\beta}_{12} &\approx -1.433, & \hat{\alpha}\hat{\beta}_{13} &= 0.4 & \hat{\alpha}\hat{\beta}_{14} &\approx 3.067
 \end{aligned}$$

which match our calculations. Estimates for other parameters can be determined by the zero-sum constraints.

Example: Sprouting Barley — ANOVA Table

```
anova(lmbarley)
Analysis of Variance Table

Response: y

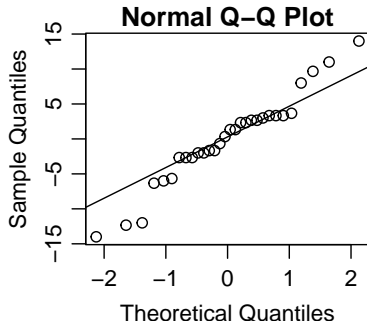
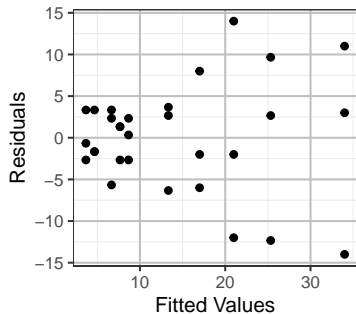
          Df Sum Sq Mean Sq F value Pr(>F)
waterfac   1  1178    1178   19.72 0.00025
weekfac    4   1321     330    5.53 0.00364
waterfac:weekfac 4    209     52    0.87 0.49673
Residuals 20   1195     60
```

Conclusion:

- ▶ It looks like both water and week main effects are significant, but their interactions are not
- ▶ Wait! Need to check model assumptions before making conclusions.

Example: Sprouting Barley — Checking Assumptions (1)

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

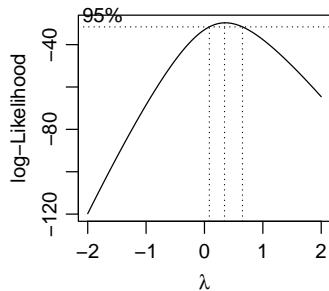


Spot any problem(s)?

- ▶ size of noise increases w/ fitted values
- ▶ normal QQ plot appear heavy-tailed

Example: Sprouting Barley — Checking Assumptions (2)

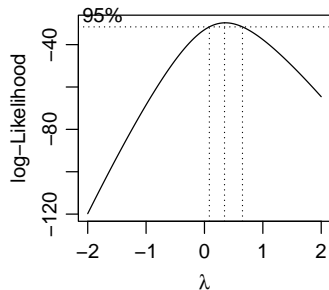
```
library(MASS)  
boxcox(lmbarley)
```



What remedy does the Box-Cox method suggest?

Example: Sprouting Barley — Checking Assumptions (2)

```
library(MASS)
boxcox(lmbarley)
```



What remedy does the Box-Cox method suggest?

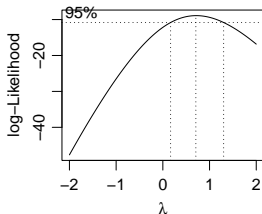
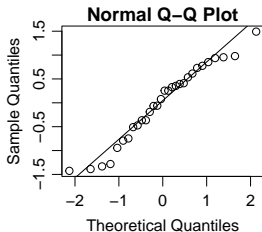
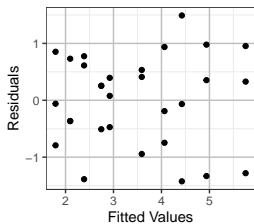
As the 95% for λ includes $1/2$, we thus take square root of the response and fit a new model below.

```
lmbarley2 = lm(sqrt(y) ~ waterfac*weekfac, data=barley)
```

Again, we should check model assumptions of the new model before interpreting the ANOVA table.

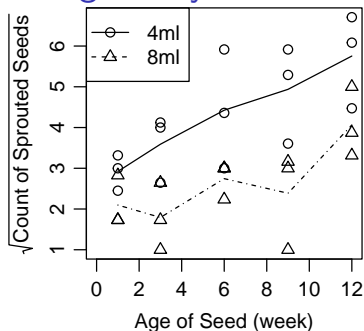
Need to Check Assumptions Again After Transformation

```
ggplot(barley, aes(x=fitted(lmbarley2), lmbarley2$res)) +  
  geom_point() + labs(x="Fitted Values", y="Residuals")  
par(mai=c(.6,.6,.25,.02),mgp=c(2,.7,0))  
qqnorm(lmbarley2$res)  
qqline(lmbarley2$res)  
boxcox(lmbarley2)
```



- ▶ size of noise appear constant, not changing with fitted values
- ▶ no apparent non-normality from the normal QQ plot
- ▶ 95% CI for λ includes 1, meaning the current response is fine, no further transformation is required.

Sprouting Barley — ANOVA Table After Transformation



```
anova(lmbarley2)
Analysis of Variance Table
```

```
Response: sqrt(y)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
waterfac	1	21.89	21.89	23.76	0.000092
weekfac	4	21.89	5.47	5.94	0.0026
waterfac:weekfac	4	2.25	0.56	0.61	0.6601
Residuals	20	18.43	0.92		

Factorial Designs v.s. One-At-a-Time Designs

When there are two factors A and B of interest, we could conduct two separate experiments and change only one factor at a time,

Experiment #1

A = 1	A = 2	A = 3	A = 4

Experiment #2

B = 1	B = 2	B = 3

rather than a two-way factorial design

	A = 1	A = 2	A = 3	A = 4
B = 1				
B = 2				
B = 3				

Advantage of Factorial Designs

Factorial designs are superior to one-at-a-time designs that change only one factor at a time because factorial design can

- ▶ test the effects of both factors at once — more *efficient* than one-at-a-time design, taking fewer experimental units to attain the same goal;
- ▶ investigate *interaction* of factors, but one-at-a-time designs cannot.
- ▶ broaden the inductive base for generalizing our results by trying a wide range of treatments