8.1-8.6 Two-Way Factorial Designs

Yibi Huang

Problem 8.1 — 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 germinating

	Age of Seeds (weeks)				
water	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, $L_2, \ldots,$ L_k levels, a **full k-way factorial design** include all the $L_1 \times L_2 \times \cdots \times L_k$ 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_{112} $\vdots$ y_{11n}	y_{121} y_{122} $\vdots$ y_{12n}	$\cdots$ $\cdots$	y_{1b1} y_{1b2} $\vdots$ y_{1bn}
A-level 2	y_{211} y_{212} $\vdots$ y_{21n}	y_{221} y_{222} $\vdots$ y_{22n}	$\cdots$ $\cdots$	y_{2b1} y_{2b2} $\vdots$ y_{2bn}
$\vdots$ $\vdots$	$\vdots$ $\vdots$	$\vdots$ $\vdots$	$\ddots$ $\ddots$	$\vdots$ $\vdots$
A-level a	y_{a11} y_{a12} $\vdots$ y_{a1n}	y_{a21} y_{a22} $\vdots$ y_{a2n}	$\cdots$ $\cdots$	y_{ab1} y_{ab2} $\vdots$ y_{abn}

Display of Data from Two Way Factorial Designs

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

	Age of Seeds (weeks)					Row means
Cell means $\bar{y}_{ij\bullet}$	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 gemination? Does the age of seeds have an effect?

Getting Cell Means, Row Means, Column Means in R

Cell means (average of the 3 values in each cell):

```
> barley = read.table("SproutingBarley.txt",header=T)
> library(mosaic)
> mean(y ~ week+ water, data=barley)
      1.4       3.4       6.4       9.4      12.4
 8.666667 13.333333 21.000000 25.333333 34.000000
      1.8       3.8       6.8       9.8      12.8
 4.666667  3.666667  7.666667  6.666667 17.000000
```

Row means:

```
> mean(y ~ water, data=barley)
        4         8
20.466667  7.933333
```

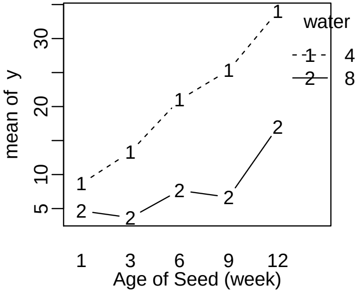
Column means:

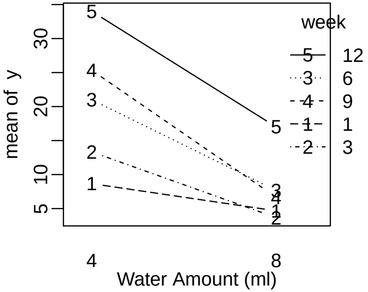
```
> mean(y ~ week, data=barley)
        1         3         6         9        12
 6.666667  8.500000 14.333333 16.000000 25.500000
```

Graphical Display of Data — Interaction Plots

cell means $\overline{y}_{ij\bullet}$	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: plotting cell means ($\overline{y}_{ij\bullet}$) against levels of one factor (A or B), with different lines for the other factor (B or A)





Means Model for a Two-Way Factorial Design

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

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

- y_{ijk} = the kth replicate in the treatment formed from the ith level of factor A and jth 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 ith level of factor A and jth level of factor B
- The means model regards the 2-way factorial design as a CRD with $a \times b$ treatments, ignoring the factorial structure of the treatments.

Main Effects (1)

Though the means model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ ignores factorial structure of the treatments, one can use appropriate *contrasts* to explore the effects of the two factors.

E.g., if one wants to compare the effects of level 1 and 2 of factor A, one can use the contrast

$$C = \frac{\mu_{11} + \mu_{12} \cdots + \mu_{1b}}{b} - \frac{\mu_{21} + \mu_{22} \cdots + \mu_{2b}}{b} = \overline{\mu}_{1\bullet} - \overline{\mu}_{2\bullet}$$

In general, if one wants to compare between levels of factor A, the contrasts are all of the form $C = \overline{\mu}_{i_1\bullet} - \overline{\mu}_{i_2\bullet}$. Observe

$$\overline{\mu}_{i\bullet} = \text{mean response at the } i\text{th level of factor A,}$$
$$\text{averaged over all levels of factor B}$$

We call $\overline{\mu}_{i\bullet}$, $i = 1, \ldots, a$, as the **main effects** of factor A.

Similarly, $\overline{\mu}_{\bullet j}$, $j = 1, \ldots, b$, are called the **main effects** of factor B.

Main Effects (2)

As $\overline{\mu}_{i\bullet}$'s and $\overline{\mu}_{\bullet j}$'s are important parameters. We thus give them new notations

$$\begin{aligned}\overline{\mu}_{i\bullet} &= \mu + \alpha_i = \text{overall effect } + \text{ effect due to factor A}\\ \overline{\mu}_{\bullet j} &= \mu + \beta_j = \text{overall effect } + \text{ effect due to factor B}\end{aligned}$$

in which, $\mu = \overline{\mu}_{\bullet\bullet}$ is the overall average of all μ_{ij}'s.

We also call α_i's as the **main effect** of factor A,
and β_j's as the **main effect** of factor B.

Observe that only $a-1$ of the α_i's can be arbitrary since

$$\begin{aligned}\sum_{i=1}^{a}\alpha_i = \sum_{i=1}^{a}\overline{\mu}_{i\bullet} - a\overline{\mu}_{\bullet\bullet} &= \sum_{i=1}^{a}\Big(\overbrace{\frac{1}{b}\sum\nolimits_{j=1}^{b}\mu_{ij}}^{\overline{\mu}_{i\bullet}}\Big) - a\overline{\mu}_{\bullet\bullet}\\ &= \frac{1}{b}\mu_{\bullet\bullet} - \frac{a}{ab}\mu_{\bullet\bullet} = 0\end{aligned}$$

Similarly, one can show that $\sum_{j=1}^{b}\beta_j = 0$.

Interaction (1)

Two factors A and B are said to have a **two-way interaction** if the effects of factor A change with the levels of factor B.

To be more specific, based on the means model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, the effect of changing factor A from level i_1 to level i_2 is

$$\begin{cases} \mu_{i_2 j_1} - \mu_{i_1 j_1} & \text{if factor B is fixed at level } j_1 \\ \mu_{i_2 j_2} - \mu_{i_1 j_2} & \text{if factor B is fixed at level } j_2 \end{cases}$$

If

$$\mu_{i_2 j_2} - \mu_{i_1 j_2} - (\mu_{i_2 j_1} - \mu_{i_1 j_1}) = \mu_{i_1 j_1} - \mu_{i_1 j_2} - \mu_{i_2 j_1} + \mu_{i_2 j_2} = 0,$$

then level (i_1, i_2) of factor A doesn't interact with level (j_1, j_2) of factor B.

If none of the levels of factor A interact the levels of factor B, i.e.,

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

then we say factor A and factor B have **no interaction**.

Interaction (2)

If factor A and factor B have no interaction, we claim that

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

in which $\mu = \overline{\mu}_{\bullet\bullet}$, $\alpha_i = \overline{\mu}_{i\bullet} - \overline{\mu}_{\bullet\bullet}$, and $\beta_i = \overline{\mu}_{\bullet j} - \overline{\mu}_{\bullet\bullet}$.

Proof.

$$\begin{aligned}
&\mu_{ij} - \mu - \alpha_i - \beta_j \\
&= \mu_{ij} - \overline{\mu}_{\bullet\bullet} + (\overline{\mu}_{\bullet\bullet} - \overline{\mu}_{i\bullet}) + (\overline{\mu}_{\bullet\bullet} - \overline{\mu}_{\bullet j}) \\
&= \mu_{ij} - \overline{\mu}_{i\bullet} - \overline{\mu}_{\bullet j} + \overline{\mu}_{\bullet\bullet} \\
&= \mu_{ij} - \frac{1}{b}\sum_{m=1}^{b}\mu_{im} - \frac{1}{a}\sum_{\ell=1}^{m}\mu_{\ell j} + \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}\mu_{\ell m} \\
&= \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}\mu_{ij} - \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}\mu_{im} - \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}\mu_{\ell j} + \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}\mu_{\ell m} \\
&= \frac{1}{ab}\sum_{\ell=1}^{a}\sum_{m=1}^{b}(\underbrace{\mu_{ij} - \mu_{im} - \mu_{\ell j} + \mu_{\ell m}}_{=0, \text{ since no interaction}}) = 0
\end{aligned}$$

Interaction (3)

In view of the result on the previous slide, we define the **interaction terms** of factor A and factor B as

$$\alpha\beta_{ij} \stackrel{\text{def}}{=} \mu_{ij} - \mu - \alpha_i - \beta_j, \quad \text{for all } i = 1, \ldots, a,\ j = 1, \ldots, b.$$

The interaction terms $\alpha\beta_{ij}$'s have the following properties

- $\alpha\beta_{ij} = 0$, for all i, j if the two factors do not interact
- $\sum_{i=1}^{a} \alpha\beta_{ij} = 0$ for all j and $\sum_{j=1}^{b} \alpha\beta_{ij} = 0$ for all i.
  In other words, the row sums and column sums of the array below are all 0

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

See the next slide for the proof.

Interaction (4)

$$\begin{aligned}\sum_{i=1}^{a} \alpha\beta_{ij} &= \sum_{i=1}^{a} (\mu_{ij} - \overline{\mu}_{i\bullet} - \overline{\mu}_{\bullet j} + \overline{\mu}_{\bullet\bullet}) \\ &= \mu_{\bullet j} - \Big(\sum_{i=1}^{a} \frac{1}{b} \mu_{i\bullet} \Big) - a\overline{\mu}_{\bullet j} + a\overline{\mu}_{\bullet\bullet} \\ &= \mu_{\bullet j} - \frac{1}{b}\mu_{\bullet\bullet} - \frac{a}{a}\mu_{\bullet j} + \frac{a}{ab}\mu_{\bullet\bullet} \\ &= 0,\end{aligned}$$

which is valid for all $j = 1, \ldots, b$.

HW today: Show that $\sum_{j=1}^{b} \alpha\beta_{ij} = 0$ for all i.

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

The main-effect-interaction model for a two-way factorial design is

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

Unlike the means model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ that there is no constraints on the parameter μ_{ij}'s, the main-effect-interaction model has several constraints

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

The means model and the main-effect-interaction model are related as follows

$$\mu = \overline{\mu}_{\bullet\bullet}, \quad \alpha_i = \overline{\mu}_{i\bullet} - \overline{\mu}_{\bullet\bullet}, \quad \beta_j = \overline{\mu}_{\bullet j} - \overline{\mu}_{\bullet\bullet}$$
$$\alpha\beta_{ij} = \mu_{ij} - \overline{\mu}_{i\bullet} - \overline{\mu}_{\bullet j} + \overline{\mu}_{\bullet\bullet}$$

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

$$
\begin{array}{|c|c|c|c|}
\hline
\mu_{11} & \mu_{12} & \cdots\cdots & \mu_{ab} \\
\hline
\mu_{21} & \mu_{22} & \cdots\cdots & \mu_{2b} \\
\hline
\vdots & \vdots & \ddots & \vdots \\
\hline
\mu_{a1} & \mu_{a2} & \cdots\cdots & \mu_{ab} \\
\hline
\end{array}
$$

$$
= \begin{array}{|c|}
\hline
\\
\mu \\
\\
\hline
\end{array}
+ \begin{array}{|c|}
\hline
\alpha_1 \\
\hline
\alpha_2 \\
\hline
\vdots \\
\hline
\alpha_a \\
\hline
\end{array}
+ \begin{array}{|c|c|c|c|}
\hline
\beta_1 & \beta_2 & \cdots\cdots & \beta_b \\
\hline
\end{array}
$$

$$
+ \begin{array}{|c|c|c|c|}
\hline
\alpha\beta_{11} & \alpha\beta_{12} & \cdots\cdots & \alpha\beta_{ab} \\
\hline
\alpha\beta_{21} & \alpha\beta_{22} & \cdots\cdots & \alpha\beta_{2b} \\
\hline
\vdots & \vdots & \ddots & \vdots \\
\hline
\alpha\beta_{a1} & \alpha\beta_{a2} & \cdots\cdots & \alpha\beta_{ab} \\
\hline
\end{array}
$$

Additive Model

A model is said to be **additive** if all interaction terms are 0.

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

In other words, additive models for two-way factorial designs assume no interactions between the two factors.

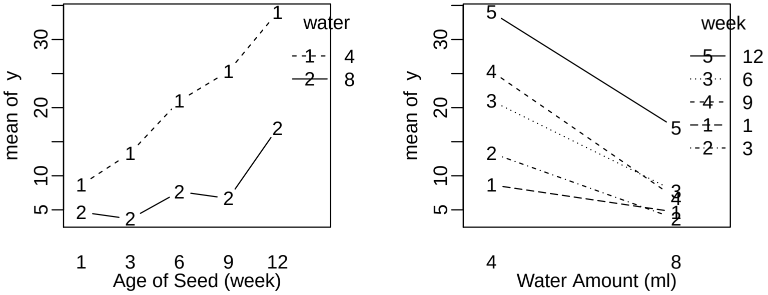
An additive model also has constraints on parameters

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0.$$

Interaction Plot Revisit

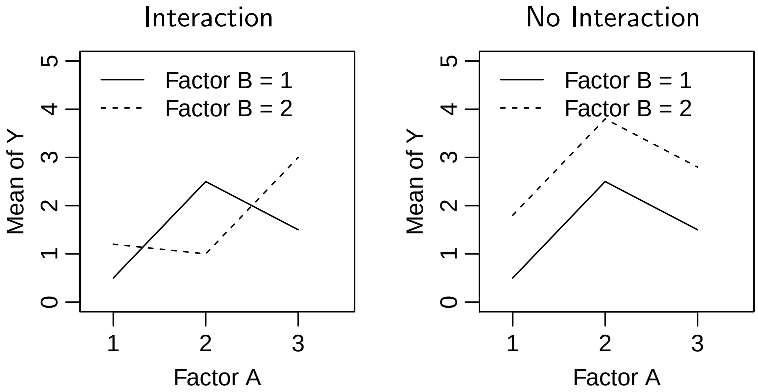
Plot cell means ($\bar{y}_{ij\bullet}$) against levels of one factor (i or j), with different lines for the other factor (j or i)

```
barley = read.table("SproutingBarley.txt",header=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)"))
```



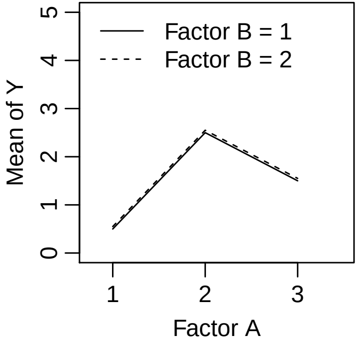
Interaction Plots Revisit(2)

Parallel lines indicate no interaction.



Interaction Plot Revisit (3)

What does the interaction plot below tell us?



Estimation of Parameters (1)

Parameter estimation in a balanced factorial design is straightforward. For the means model and the effects model,

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

the parameter estimates are

$$\begin{aligned}
\widehat{\mu}_{ij} &= \overline{y}_{ij\bullet} \\
\widehat{\mu} &= \overline{y}_{\bullet\bullet\bullet}, \\
\widehat{\alpha}_i &= \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet}, \\
\widehat{\beta}_j &= \overline{y}_{\bullet j\bullet} - \overline{y}_{\bullet\bullet\bullet} \\
\widehat{\alpha\beta}_{ij} &= \overline{y}_{ij\bullet} - \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet j\bullet} + \overline{y}_{\bullet\bullet\bullet}
\end{aligned}$$

Observe the estimates also satisfy the zero-sum constraints:

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

Estimation of Parameters (2)

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

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

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

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

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

Recall in a regression model, the estimate of a coefficient will change with the model formula.

Fitted Values for a Main-Effect-Interaction Model

For a main-effect-interaction model, the fitted value for y_{ijk} is simply the cell mean $\overline{y}_{ij\bullet}$ because

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

Fitted Values for an Additive Model

For an additive model (**no interaction**), the fitted value for y_{ijk} is

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

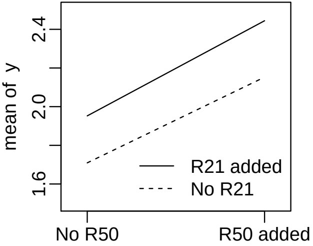
Example 8.6 Bacteria in Cheese (p.178 in Oehlert)

- 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	2.211
	1.601	1.673
	1.830	1.973
R50 added	2.032	2.091
	2.017	2.255
	2.409	2.987

$\Rightarrow$

	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?

Example 8.6 Bacteria in Cheese (p.178 in Oehlert)

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

$$\begin{aligned}
\widehat{\mu} &= \overline{y}_{\bullet\bullet\bullet} = 2.065 \\
\widehat{\alpha}_1 &= \overline{y}_{1\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet} = 1.831 - 2.065 = -0.234 \\
\widehat{\beta}_1 &= \overline{y}_{\bullet 1\bullet} - \overline{y}_{\bullet\bullet\bullet} = 1.931 - 2.065 = -0.134 \\
\widehat{\alpha\beta}_{11} &= \overline{y}_{11\bullet} - \overline{y}_{1\bullet\bullet} - \overline{y}_{\bullet 1\bullet} + \overline{y}_{\bullet\bullet\bullet} \\
&= 1.709 - 1.831 - 1.931 + 2.065 = 0.012
\end{aligned}$$

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

$$\begin{aligned}
\widehat{\alpha}_1 + \widehat{\alpha}_2 = 0 \quad &\Rightarrow \quad \widehat{\alpha}_2 = -\widehat{\alpha}_1 = 0.234 \\
\widehat{\beta}_1 + \widehat{\beta}_2 = 0 \quad &\Rightarrow \quad \widehat{\beta}_2 = -\widehat{\beta}_1 = 0.134 \\
\widehat{\alpha\beta}_{11} + \widehat{\alpha\beta}_{12} = 0 \quad &\Rightarrow \quad \widehat{\alpha\beta}_{12} = -\widehat{\alpha\beta}_{11} = -0.012 \\
\widehat{\alpha\beta}_{11} + \widehat{\alpha\beta}_{21} = 0 \quad &\Rightarrow \quad \widehat{\alpha\beta}_{21} = -\widehat{\alpha\beta}_{11} = -0.012 \\
\widehat{\alpha\beta}_{12} + \widehat{\alpha\beta}_{22} = 0 \quad &\Rightarrow \quad \widehat{\alpha\beta}_{22} = -\widehat{\alpha\beta}_{12} = 0.012
\end{aligned}$$

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 is still valid.

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

where

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

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

$$\begin{aligned}
\text{d.f. for SST} &= \text{total \# of observations} - 1 = abn - 1\\
\text{d.f. for SS}_{trt} &= \text{\# of treatments} - 1 = ab - 1\\
\text{d.f. for SSE} &= \text{total \# of observations} - \text{\# of treatments}\\
&= abn - ab = ab(n-1)
\end{aligned}$$

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	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$a-1$
SS_B	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{\bullet j\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$b-1$
SS_{AB}	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{ij\bullet} - \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet j\bullet} + \overline{y}_{\bullet\bullet\bullet})^2$	$(a-1)(b-1)$
SS_{trt}	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{ij\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$ab-1$

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

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

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

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

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

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

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

Expected Values for the Mean Squares

Just like CRD, the mean squares for factorial design are the sum of squares divided by the corresponding d.f.

$$\mathrm{MS}_A = \frac{\mathrm{SS}_A}{a-1},\ \mathrm{MS}_B = \frac{\mathrm{SS}_B}{b-1},\ \mathrm{MS}_{AB} = \frac{\mathrm{SS}_{AB}}{(a-1)(b-1)},\ \mathrm{MSE} = \frac{\mathrm{SSE}}{ab(n-1)}.$$

Under the effects model for a balanced two-way factorial,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad \varepsilon_{ijk}\text{'s are i.i.d. } N(0, \sigma^2).$$

one can show that

$$\mathbb{E}(\mathrm{MS}_A) = \sigma^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2, \quad \mathbb{E}(\mathrm{MS}_B) = \sigma^2 + \frac{an}{b-1}\sum_{j=1}^{b}\beta_j^2$$

$$\mathbb{E}(\mathrm{MS}_{AB}) = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha\beta_{ij}^2, \quad \mathbb{E}(\mathrm{MSE}) = \sigma^2$$

Again the MSE is an unbiased estimator of σ^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?

   $$\begin{cases} H_0 : \alpha_1 = \cdots = \alpha_a = 0 \\ H_a : \text{not all } \alpha_i\text{'s} = 0, \end{cases} \Rightarrow F_A = \frac{\text{MS}_A}{\text{MSE}} \sim F_{a-1,\, ab(n-1)} \text{ under H}_0.$$

2. Does factor B has an effect on the response?
   E.g. does the water amount has an effect on germination?

   $$\begin{cases} H_0 : \beta_1 = \cdots = \beta_b = 0 \\ H_a : \text{not all } \beta_i\text{'s} = 0, \end{cases} \Rightarrow F_B = \frac{\text{MS}_B}{\text{MSE}} \sim F_{b-1,\, ab(n-1)} \text{ under H}_0.$$

3. Does the effect of factor A interact with that of factor B?
   E.g., does the effect of age change with water amount?

   $$\begin{cases} H_0 : \alpha\beta_{ij} = 0 \text{ for all } i, j \\ H_a : \alpha\beta_{ij} \neq 0 \text{ for some } i, j \end{cases} \Rightarrow F_{AB} = \frac{\text{MS}_{AB}}{\text{MSE}} \sim F_{(a-1)(b-1),\, ab(n-1)} \text{ under H}_0.$$

Example 8.6 Bacteria in Cheese (p.178 in Oehlert)

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

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

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

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

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

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 8.6 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 get the ANOVA table in R as follows.

```
> 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
```

Advantage and Disadvantage of Factorial Designs

Advantage: Factorial design is 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.

Disadvantage:

- If there are many factors or many levels, the size of the experiment can be very large.
  Remedy: fractional factorial designs (Chapter 18)