Lab 2: Factorial Experiments

M. George Akritas
Factorial Experiments: Main Effects and Interactions
A statistical experiment involving several factors is called a *factorial experiment* if all factor-level combinations are considered.

For example, the two-factor study portrayed below is a factorial experiment if all 8 treatments are included in the study.

<table>
<thead>
<tr>
<th>Factor A</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor B</td>
<td>1</td>
<td>Tr11</td>
<td>Tr12</td>
<td>Tr13</td>
</tr>
<tr>
<td>1</td>
<td>Tr21</td>
<td>Tr22</td>
<td>Tr23</td>
<td>Tr24</td>
</tr>
</tbody>
</table>

**Figure:** Treatments, or factor-level combinations, in a two-factor study
In product and industrial design it is typical to consider the impact of a large number of factors on a number of quality characteristics of a product.

For example, in car manufacturing quality aspects range from the car’s handling to the door’s holding ability for remaining open when the car is parked uphill. Optimization of such quality characteristics is only possible through factorial experimentation.
How does one compare the levels of a factor in the presence of another factor?

Comparative boxplots for the levels of each factor individual factor fail to capture possible synergistic effects among the levels of different factors.

Such synergistic effects, called interactions, may result in some factor level combinations yielding improved or diminished response levels far beyond what can be explained by any differences between the levels of the individual factors.

Pinot Gris goes well with abalone, while Chablis Premier Cru is better for enhancing the taste of oysters (cf. Hugh Johnson’s Wine Book); thus, the two levels of the factor ”white wine” interact with the two levels of the factor ”appetizer”. Similarly different spices interact with different types of foods.
1. In agriculture, different types of fertilization may interact with different types of soil as well as different levels of watering.

2. The June 2008 issue of Development features research suggesting interaction between two transcription factors that regulate the development and survival of retinal ganglion cells.

3. In QoS, an IEEE article (Selected Areas in Communications, Vol. 22, pp. 1335-1346), considers the impact of several factors on total throughput and average delay as measures of service delivery. Due to interactions, the article concludes, the factors cannot be studied in isolation.
Example (Bio-fuel corn)

<table>
<thead>
<tr>
<th></th>
<th>Fertilizer</th>
<th>Row Averages</th>
<th>Main Row Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>Seed A</td>
<td>$\mu_{11} = 107$</td>
<td>$\mu_{12} = 111$</td>
<td>$\mu_1. = 109$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_1 = -0.25$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed B</td>
<td>$\mu_{21} = 109$</td>
<td>$\mu_{22} = 110$</td>
<td>$\mu_2. = 109.5$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2 = 0.25$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column Averages</td>
<td>$\mu_1 = 108$</td>
<td>$\mu_2 = 110.5$</td>
<td>$\mu_{..} = 109.25$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main Column Effects</td>
<td>$\beta_1 = -1.25$</td>
<td>$\beta_2 = 1.25$</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure:** A $2 \times 2$ design with interaction (non-additive design)
Definition

When a change in the level of factor A has different effects on the levels of factor B we say that there is interaction between the two factors. The absence of interaction is called additivity.

Under additivity there is an indisputably best level for each factor and the best factor-level combination is that of the best level of factor A with the best level of factor B. See the following variation of the bio-fuel and fertilizer example.
### Factorial Experiments: Main Effects and Interactions

#### Fertilizer

<table>
<thead>
<tr>
<th></th>
<th>Fertilizer</th>
<th>Row Averages</th>
<th>Main Row Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>( \mu_{11} = 107 )</td>
<td>( \bar{\mu}_1. = 109 )</td>
<td>( \alpha_1 = -1 )</td>
</tr>
<tr>
<td>II</td>
<td>( \mu_{12} = 111 )</td>
<td>( \bar{\mu}_1 \cdot = 109 )</td>
<td>( \alpha_2 = 1 )</td>
</tr>
</tbody>
</table>

#### Seed A

- \( \mu_{21} = 109 \)
- \( \mu_{22} = 113 \)

#### Seed B

- \( \mu_{21} = 109 \)
- \( \mu_{22} = 113 \)

#### Column Averages

- \( \bar{\mu}_1 = 108 \)
- \( \bar{\mu}_2 = 112 \)
- \( \bar{\mu}_1 \cdot = 110 \)
- \( \bar{\mu}_2 \cdot = 111 \)
- \( \bar{\mu}_1 \cdot \cdot = 110 \)
- \( \bar{\mu}_2 \cdot \cdot = 110 \)

#### Main Column Effects

- \( \beta_1 = -2 \)
- \( \beta_2 = 2 \)
- \( \beta_1 \cdot = 0 \)
- \( \beta_2 \cdot = 0 \)

**Figure:** A 2 x 2 design with no interaction (additive design)
Under additivity, the comparison of the levels of each factor is typically based on the so-called **main effects**. The *main row effects*, denoted by $\alpha_i$, and *main column effects*, denoted by $\beta_j$, are defined as

$$
\alpha_i = \mu_i - \mu_{..}, \quad \beta_j = \mu_{.j} - \mu_{..}
$$

Moreover, the cell means are given in an additive manner in terms of the main effects:

$$
\mu_{ij} = \mu_{..} + \alpha_i + \beta_j
$$
When there is interaction between the two factors, the cell means are not given by the additive relation (2.2). The discrepancy/difference between the left and right hand sides of this relation quantifies the interaction effects:

$$\gamma_{ij} = \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j) \quad \text{Interaction Effects}$$

(2.3)
Example

Compute the interaction effects in the design of Figure 2.

Solution. Using the information shown in Figure 2, we have

\[ \gamma_{11} = \mu_{11} - \overline{\mu ..} - \alpha_1 - \beta_1 = 107 - 109.25 + 0.25 + 1.25 = -0.75 \]

\[ \gamma_{12} = \mu_{12} - \overline{\mu ..} - \alpha_1 - \beta_2 = 111 - 109.25 + 0.25 - 1.25 = 0.75 \]

\[ \gamma_{21} = \mu_{21} - \overline{\mu ..} - \alpha_2 - \beta_1 = 109 - 109.25 - 0.25 + 1.25 = 0.75 \]

\[ \gamma_{22} = \mu_{22} - \overline{\mu ..} - \alpha_2 - \beta_2 = 110 - 109.25 - 0.25 - 1.25 = -0.75. \]
Data from a two-factor factorial experiment are typically denoted using three subscripts as shown below.

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>$x_{11k}$, $k = 1, \ldots, n_{11}$</td>
</tr>
<tr>
<td></td>
<td>$x_{12k}$, $k = 1, \ldots, n_{12}$</td>
</tr>
<tr>
<td></td>
<td>$x_{13k}$, $k = 1, \ldots, n_{13}$</td>
</tr>
<tr>
<td></td>
<td>$x_{14k}$, $k = 1, \ldots, n_{14}$</td>
</tr>
<tr>
<td>2</td>
<td>$x_{21k}$, $k = 1, \ldots, n_{21}$</td>
</tr>
<tr>
<td></td>
<td>$x_{22k}$, $k = 1, \ldots, n_{22}$</td>
</tr>
<tr>
<td></td>
<td>$x_{23k}$, $k = 1, \ldots, n_{23}$</td>
</tr>
<tr>
<td></td>
<td>$x_{24k}$, $k = 1, \ldots, n_{24}$</td>
</tr>
</tbody>
</table>

**Figure:** Data notation in a $2 \times 4$ factorial experiment
Sample versions of the main effects and interactions are defined similarly using the cell means

\[
\bar{x}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk},
\]

Sample Mean of Observations in Cell \((i, j)\) (2.4)

instead of the population means \(\mu_{ij}\):

\[
\hat{\alpha}_i = \bar{x}_i - \bar{x}_{..}, \quad \hat{\beta}_j = \bar{x}_j - \bar{x}_{..}
\]

Sample Main Row and Column Effects (2.5)

\[
\hat{\gamma}_{ij} = \bar{x}_{ij} - \left( \bar{x}_{..} + \hat{\alpha}_i + \hat{\beta}_j \right)
\]

Sample Interaction Effects (2.6)
The computation of the sample main row effects for a $3 \times 4$ factorial experiment is illustrated schematically below.

<table>
<thead>
<tr>
<th>Row Factor</th>
<th>Column Factor</th>
<th>Row Averages</th>
<th>Main Row Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\bar{x}_{11}$</td>
<td>$\bar{x}_{12}$</td>
<td>$\bar{x}_{13}$</td>
</tr>
<tr>
<td>2</td>
<td>$\bar{x}_{21}$</td>
<td>$\bar{x}_{22}$</td>
<td>$\bar{x}_{23}$</td>
</tr>
<tr>
<td>3</td>
<td>$\bar{x}_{31}$</td>
<td>$\bar{x}_{32}$</td>
<td>$\bar{x}_{33}$</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}_{..}$</td>
<td>$\bar{x}_{..}$</td>
<td>$\bar{x}_{..}$</td>
</tr>
</tbody>
</table>

**Figure:** Calculation of the sample main row effects in a $3 \times 4$ design
Main Effects and Interactions in R

Example (Cloud seeding in Tasmania)

Use R to compute the cell means, and the main and interaction effects for the factors seed and season using the rainfall data in the site given below (source: Miller, A.J, et al. (1979), Analyzing the results of a cloud-seeding experiment in Tasmania, *Communications in Statistics - Theory & Methods*, A8(10), 1017-1047. ¹)

Solution. First import the data in R using

cs=read.table("http://sites.stat.psu.edu/~mga/401/Data/CloudSeed2w.txt", header=T)

¹See also the related article by Morrison,A.E, et al. (2009), On the Analysis of a Cloud Seeding Dataset over Tasmania, *American Meteorological Society*, 48, 1267-1280
Main Effects and Interactions in R

and then use the following commands:

```r
mcm = tapply(cs$rain, cs[,c(2,3)], mean)  # the matrix of cell means
alphas = rowMeans(mcm) - mean(mcm)  # the vector of main row effects
betas = colMeans(mcm) - mean(mcm)  # the vector main column effects
gammas = t(t(mcm-mean(mcm)-alphas)-betas)  # the matrix of interaction effects.
```
As we did before, we stress again that the sample versions of the main effects and interactions approximate but, in general, they are not equal to their population counterparts.

In particular, the sample interaction effects will not be equal to zero even if the design is additive.

The interaction plot is a useful graphical technique for assessing whether the sample interaction effects are sufficiently different from zero to imply a non-additive design.

For each level of one factor, say factor B, the interaction plot traces the cell means along the levels of the other factor. If the design is additive, these traces (also called profiles) should be approximately parallel.
attach(cs) # so variables can be referred to by name

interaction.plot(season, seed, rain, col=c(2,3), lty = 1, xlab="Season", ylab="Cell Means of Rainfall", trace.label="Seeding")
Figure: interaction plot for the rainfall data