DOE
DESIGN OF EXPERIMENT
EXPERIMENTAL DESIGN

Lesson 3
Framework of lesson 3:

• Two factor Project
• Analysis of Variance
• Principal effects and interaction
• What is interaction and how can we represent it graphically
• How to act in case of absence of repetitive tests
• Practical examples
TWO FACTORS PROJECT

• Until now it has been taken care of plans with a single factor. It was supposed to want to assess the impact of the brand of a battery (X, the independent variable) on the duration (Y, the dependent variable). Suppose you now want to investigate the impact of another factor: the instrument, which must be powered by the battery.

• Most obvious possibility: to run two distinct monofactorial plans.

• Advantages: easy plan of the experiments and data analysis.

• Disadvantages: two distinct project could lead to erroneous conclusions.
TWO FACTORS PROJECT

• Considering an instrument, one specific brand could be the best, but, considering another instrument, another brand could allow longer battery life.

• This synergic, or interactive, effect between two factors is called “interaction”

• Due to this effect can happen that, if the optimal level of factor A (with fixed factor B) is \( A_0 \), and if the optimal level of B (with fixed A) is \( B_1 \), the combination \( A_0, B_1 \) could not be the optimal one.
TWO FACTORS PROJECT

• We study the impact of the two factors over the answer.

• From the former example:

• $X_1 = \text{Brand}, X_2 = \text{Instrument}, Y = \text{battery life}$

<table>
<thead>
<tr>
<th>Instrument ($X_2$)</th>
<th>Brand ($X_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>17.9; 18.1</td>
</tr>
<tr>
<td>2</td>
<td>18.0; 18.2</td>
</tr>
<tr>
<td>3</td>
<td>18.0; 17.8</td>
</tr>
</tbody>
</table>
TWO FACTORS PROJECT

• Factor Brand ($X_1$): 4 levels $\rightarrow$ 4 columns ($C = 4$)
• Factor Instrument ($X_2$): 3 levels $\rightarrow$ 3 rows ($R = 3$)
• Repetitions : 2 $\rightarrow$ 2 elements for each cell $\rightarrow$ $n = 2$ ($n_{ij}$)
• As a consequence in the matrix we have $R \cdot C$ cells, each with $n$ values. Therefore, the results are $RCn$
STATISTICAL MODEL

- $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$
- $Y_{ijk} = \mu + \rho_i + \tau_j + l_{ij} + \varepsilon_{ijk}$

- $i = 1, \ldots, R$ (number of rows)
- $j = 1, \ldots, C$ (number of columns)
- $k = 1, \ldots, n$ (number of repetitions)
STATISTICAL MODEL

- \( Y_{ijk} = \mu + \rho_i + \tau_j + I_{ij} + \varepsilon_{ijk} \)

- \( Y_{ijk} = \) value of the i-th row, j-th columns, k-th repetition.
- \( \mu = \) Real value of the Grand Mean
- \( \rho_i = \) Real deviation of the i-th row mean from the Grand Mean
- \( \tau_j = \) Real deviation of the j-th columns from the Grand Mean
- \( I_{ij} = \) measure of the interaction associated to the i-th row and the j-th column
- \( \varepsilon_{ijk} = \) error or noise generated from the experimental uncertainty, difference between each experimental value (dependend on i, j, k) and the real mean of the cell.
STATISTICAL MODEL

- $\bar{Y}_{...}$ = mean over all the values $\rightarrow$ Estimator of $\mu$
- $\bar{Y}_{i..}$ = mean computed over the values of the $i$-th row
- $\bar{Y}_{..j}$ = mean computed over the values of the $j$-th column
- $\bar{Y}_{ij.}$ = mean computed over the values of the $(i,j)$-th cell

$Y_{ijk} = \mu + \rho_i + \tau_j + l_{ij} + \varepsilon_{ijk}$

$Y_{ijk} = \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{..j} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...}) + (\bar{Y}_{ijk} - \bar{Y}_{ij.})$
STATISTICAL MODEL

\[ Y_{ijk} = Y_{..} \]

- Estimator of \( \mu \)
- Estimator of \( \rho_i \)
- Estimator of \( \tau_j \)
- Estimator of \( l_{ij} \)
- Estimator of \( \varepsilon_{ijk} \)

All the terms are intuitive, with the exception for the one of Interaction:

\[ (Y_{ij} - Y_{i..} - Y_{.j} + Y_{..}) \]
The interaction term can be rewritten in a more intuitive way:

\[
(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...}) =
\]

\[
= (\bar{Y}_{ij} - \bar{Y}_{...}) - (\bar{Y}_{i..} - \bar{Y}_{...}) - (\bar{Y}_{.j} - \bar{Y}_{...})
\]

How much a cell differs from the Grand Mean

Difference between the i-th row and the Grand Mean ($\rho_i$)

Difference between the j-th column and the Grand Mean ($\tau_j$)

Adjustment for the “row effect”

Adjustment for the “column effect”
INTERACTION

• We suppose a negligible error, we averaged over infinite tests; \( Y_{ij} \equiv \mu_{ij} \), real value of the cell.
• The value of the previous term deducts by the deviation of a cell the effects of row and column, but is in general different from zero, because it takes over an additional effect: the interaction.
INTERACTION

- If $A_H B_H = 13$, no interaction
- Se $A_H B_H > 13$, positive interaction
- Se $A_H B_H < 13$, negative interaction

- B factor constant: $B = B_L$
- If factor A moves from $A_L$ a $A_H$, the result increase from 5 to 10 (+5)

- A factor constant : $A = A_L$
- If the factor B moves from $B_L$ to $B_H$, the result increase from 5 to 8 (+3)

- In the absence of interaction the two effects should be added:
- Therefore moving from $A_L B_L$ to $A_H B_H$, the result should increase of 8 (5+3), changing from 5 to 13.
INTERACTION

• Practical case: let’s suppose to run a statistical analysis with the aim to evaluate the impact over the sales of a product related to two possible marketing actions: increase the space on the shelf or implement a discount.

• Let’s suppose that, in absence of a discount, the increase of space make increase the sales of the 8%, and that, without increasing the space, the discount make increase the sales of the 12%.

• If at the same time we increase the space and apply a discount sales should increase of the 20%. If this is the case (sum of effects), there is no interaction. If sales increase more that the 20%, we have a positive interaction, if they increase less, we have a negative interaction.

• The interaction can be defined like the degree of difference of the sum of the effects of each factor separately.
INTERACTION

• Let’s suppose that $A_H B_H = 13$:

• Looking at the definition:

$$
\bar{Y}_{HH.} = 13 \quad \bar{Y}_{H..} = 11,5 \quad \bar{Y}_{.H.} = 10,5 \quad \bar{Y}_{...} = 9
$$

$$
\bar{Y}_{HH.} - \bar{Y}_{...} = 4 \quad \bar{Y}_{H..} - \bar{Y}_{...} = 2,5 \quad \bar{Y}_{.H.} - \bar{Y}_{...} = 1,5
$$

$$
I_{ij} = (\bar{Y}_{ij} - \bar{Y}_{...}) - (\bar{Y}_{i..} - \bar{Y}_{...}) - (\bar{Y}_{.j.} - \bar{Y}_{...}) = 4 - 2,5 - 1,5 = 0
$$

<table>
<thead>
<tr>
<th></th>
<th>$B_L$</th>
<th>$B_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_L$</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>$A_H$</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

Null interaction
INTERACTION

• Let’s suppose that $A_H B_H = 17$:

• Looking at the definition:

\[
\bar{Y}_{HH} = 17 \quad \bar{Y}_{H.} = 13.5 \quad \bar{Y}_{.H} = 12.5 \quad \bar{Y} = 10
\]

\[
\bar{Y}_{HH} - \bar{Y} = 7 \quad \bar{Y}_{H.} - \bar{Y} = 3.5 \quad \bar{Y}_{.H} - \bar{Y} = 2.5
\]

\[
I_{ij} = (\bar{Y}_{ij} - \bar{Y}) - (\bar{Y}_{i.} - \bar{Y}) - (\bar{Y}_{.j} - \bar{Y}) = 7 - 3.5 - 2.5 = 1 > 0
\]

<table>
<thead>
<tr>
<th>$A_L$</th>
<th>$B_L$</th>
<th>$B_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Positive interaction
INTERACTION

• Once B factor is fixed at the low level ($B = B_L$): if the factor $A$ changes from $A_L$ to $A_H$, the result increase from 5 to 10 (+5)

• If now we fix the factor $B$ at the high level ($B = B_H$): changing $A$ from $A_L$ to $A_H$, the result increase from 8 to 17 (+9)

• We can see that changing the level of one factor, changes also the effect of the other factor. Since, in this case, the effect increase, we have a positive interaction.

• Alternative definition of interaction:
We have an interaction if the effect of one factor changes with the changes of levels of the other factor.

<table>
<thead>
<tr>
<th></th>
<th>$B_L$</th>
<th>$B_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_L$</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>$A_H$</td>
<td>10</td>
<td>17</td>
</tr>
</tbody>
</table>
INTERACTION

A different way to represent the interaction is the graphic one: null interaction (parallel lines)
A different way to represent the interaction is the graphic one: positive interaction.
A different way to represent the interaction is the graphic one: negative interaction with strong distortion.
SUM OF SQUARES

• Going back to the statistical model already shown:

\[ Y_{ijk} - \bar{Y} ... = (\bar{Y}_{i..} - \bar{Y}...) + (\bar{Y}_{.j.} - \bar{Y}...) + \\
+ (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...) + (\bar{Y}_{ijk} - \bar{Y}_{ij.}) \]

• Squaring both sides of the equation and summing (both sides) on i, j, k, we observe that all double products are deleted.
SUM OF SQUARES

\[ \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + \]

- Only i appears
- Only j appears
- Only i, j appear

- \( C = \) number of columns
  - (levels of factor 1)
- \( R = \) number of rows
  - (levels of factor 2)
- \( n = \) number of repetitions
SUM OF SQUARES

\[
\sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{..})^2 = n \cdot C \cdot \sum_{i=1}^{R} (Y_{i..} - \bar{Y}_{..})^2 + \\
+ n \cdot R \cdot \sum_{j=1}^{C} (Y_{.j.} - \bar{Y}_{..})^2 + \\
+ n \cdot \sum_{i=1}^{R} \sum_{j=1}^{C} (Y_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{..})^2 + \\
+ \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (\bar{Y}_{ijk} - \bar{Y}_{ij.})^2
\]

• C = number of columns
  – (levels of factor 1)
• R = number of rows
  – (levels of factor 2)
• n = number of repetitions
SUM OF SQUARES

\[ TSS = SSB_R + SSB_C + SSI_{R,C} + SSW \]

- With the following DoF:
  - TSS → RCn – 1 (total of experimental data -1)
  - SSB_R → R – 1 (number of rows - 1)
  - SSB_C → C – 1 (number of columns - 1)
  - SSW → RC(n-1) (in each cell there are n values, therefore each cell has (n-1) DoF, cells are RC)
  - SSI_{R,C} → (R - 1)(C - 1) (is the product of the DoF of the interactive factors). By subtraction:
    \[ RCn – 1 – (C – 1) – (R – 1) – RC(n – 1) = RCn – 1 – C + 1 – R + 1 - RCn + RC = \]
    \[ = C – R +1 + RC = (R – 1)(C – 1) \]
SUM OF SQUARES

• $SS_{R,C} \rightarrow (R - 1)(C - 1)$ (is the product of the DoF of interacting factors).

• It is possible to show it with the following example. Let’s consider a plane with two factors, 3 levels per row factor and 4 levels per column factor. Knowing the mean of each cell $Y_{ij}$ here reported and knowing all the mean of rows and columns, all the other values are automatically determined.

\[
\begin{array}{cccc}
\bar{Y}_{11.} & \bar{Y}_{12.} & \bar{Y}_{13.} & \bar{Y}_{1..} \\
\bar{Y}_{21.} & \bar{Y}_{22.} & \bar{Y}_{23.} & \bar{Y}_{2..} \\
\bar{Y}_{31.} & \bar{Y}_{32.} & \bar{Y}_{33.} & \bar{Y}_{3..} \\
\bar{Y}_{.1} & \bar{Y}_{.2} & \bar{Y}_{.3} & \bar{Y}_{.4} \\
\end{array}
\]
EXAMPLE

• Going back to the former example...

\[
\begin{array}{cccc}
\bar{Y}_{ij} & \text{Brand (X}_1) & \bar{Y}_{i..} \\
\hline
\text{Instrument (X}_2) & 1 & 2 & 3 & 4 \\
1 & 17,9; 18,1; 18,00 & 17,8; 17,8; 17,8 & 18,1; 18,2; 18,15 & 17,8; 17,9; 17,85 & 17,95 \\
2 & 18,0; 18,2; 18,1 & 18,0; 18,3; 18,15 & 18,4; 18,1; 18,25 & 18,1; 18,5; 18,25 & 18,2 \\
3 & 18,0; 17,8 17,90 & 17,8; 18,0 17,9 & 18,1; 18,3 18,20 & 18,1; 17,9 18,00 & 18,00 \\
\hline
\bar{Y}_{..j} & 18,00 & 17,95 & 18,20 & 18,05 & 18,05 \\
\bar{Y}_{..} & 18,00 & 17,95 & 18,20 & 18,05 & 18,05 \\
\end{array}
\]
EXAMPLE

$$SSB_R = 2 \cdot 4[(17,95-18,05)^2 + (18,20-18,05)^2 + (18,00-18,05)^2] =$$
$$= 8 \cdot [0,01+0,0225+0,0025]^2 = 0,28$$

$$SSB_C = 2 \cdot 3[(18,00-18,05)^2 + (17,95-18,05)^2 + (18,20-18,05)^2 +$$
$$+(18,05-18,05)^2] = 6 \cdot [0,0025+0,001+0,0225+0]^2 = 0,21$$

$$SSW = (17,90-18,00)^2 + (18,10-18,00)^2 + (17,80-17,80)^2 +$$
$$+(17,80-17,80)^2 + ... + (18,10-18,00)^2 + (17,90-18,00)^2 = 0,30$$

$$TSS = (17,90-18,05)^2 + (18,10-18,05)^2 + (17,80-18,05)^2 +$$
$$+(17,80-18,05)^2 + ... + (18,10-18,05)^2 + (17,90-18,05)^2 = 0,90$$

$$SSI_{R,C} = 0,90 – 0,28 – 0,21 – 0,30 = 0,11 (by difference)$$
## Example

<table>
<thead>
<tr>
<th>Variance</th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
<th>$F_{\text{calc}}$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SSB_R$</td>
<td>0.28</td>
<td>2</td>
<td>0.14</td>
<td>5.6</td>
<td>1.9%</td>
</tr>
<tr>
<td>$SSB_C$</td>
<td>0.21</td>
<td>3</td>
<td>0.07</td>
<td>2.8</td>
<td>8.5%</td>
</tr>
<tr>
<td>$SSI_{R,C}$</td>
<td>0.11</td>
<td>6</td>
<td>0.018</td>
<td>0.73</td>
<td>50%</td>
</tr>
<tr>
<td>$SSW$</td>
<td>0.30</td>
<td>12</td>
<td>0.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.9</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1) $H_0$: The *true means* of the rows are not different  
$H_1$: Not all the *true means* are not different  
p-value = 1.9% < 5%, reject $H_0$

2) $H_0$: The *true means* of the columns are not different  
$H_1$: Not all the *true means of the columns* are not different  
p-value = 8.5% > 5%, we accept $H_0$

3) $H_0$: There is no interaction  
$H_1$: There is interaction  
p-value = 50% > 5%, we accept $H_0$
We have the following expected values:

- $E(MSI) = \sigma^2 + V_{\text{interaction}}$; $E(MSW) = \sigma^2$
- Since $V_{\text{interaction}}$ cannot be negative and that $MSI = 0.018 < MSW = 0.025$, we get clear evidence that $V_{\text{interaction}} = 0$. Under this hypothesis we have that the expected value of MSI, $E(MSI)$, is equal to $\sigma^2$ and therefore contributes to the estimation of the error.
- In this case some texts suggest to include it into the error. We have therefore:

<table>
<thead>
<tr>
<th></th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>0.11</td>
<td>6</td>
<td>0.018</td>
</tr>
<tr>
<td>Error</td>
<td>0.30</td>
<td>12</td>
<td>0.025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>0.41</td>
<td>18</td>
<td>0.023</td>
</tr>
</tbody>
</table>
A model is called fix if both his factors are fixed (fixed level), is called random if both have random levels and it is called mixed in the intermediate situation of a factor with fixed level and one with random level.

Up to now we assumed that the model was “fixed”.

The threshold between fixed level and random level factors is very subtle.

Strict definition:

Fixed: All possible levels of the factors of which we want to study the influence are considered (Typical situation: factor sex at birth, we consider the only two possible levels, male or female)

Random: the levels are some of a wide range
**FIXED/RANDOM MODEL**

- Flexible definition:
- Fixed: we have a defined number of levels, within which we want to carry out the study. In case of a continuous variable, we consider some values of the same, for instance a set of equally spaced values.
- Random: The levels are randomly chosen among an infinite population and often they cannot be directly selected by the researcher. For instance, for the variable “rainfall”, the mm of rain fallen in one year.
FIXED/RANDOM MODEL

• Often it is difficult to decide if we are facing fixed or random level factors, and therefore fixed, random or mixed models.

• Some authors suggest that, in case of doubt, we should always assume the fixed model. In any case the discussion become irrelevant in the cases in which we can assume the absence of interaction between the factors.
FIXED/RANDOM MODEL

We just introduced the possibility for a model to be also “random” or “mixed”. In that case the expected values can be different and we should therefore modify the kind of analysis to perform for the evaluation of the effects.

<table>
<thead>
<tr>
<th></th>
<th>Fixed</th>
<th>Random</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MSB_R$</td>
<td>$\sigma^2+V_R$</td>
<td>$\sigma^2+V_R+V_I$</td>
<td>$\sigma^2+V_R$</td>
</tr>
<tr>
<td>$MSB_C$</td>
<td>$\sigma^2+V_C$</td>
<td>$\sigma^2+V_C+V_I$</td>
<td>$\sigma^2+V_C+V_I$</td>
</tr>
<tr>
<td>$MSB_I$</td>
<td>$\sigma^2 +V_I$</td>
<td>$\sigma^2 +V_I$</td>
<td>$\sigma^2+V_I$</td>
</tr>
<tr>
<td>$MSW$</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
Varying the expected values depending on the kind of model, the ratios for the identification of the row, column and interaction effect change too.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Fixed</th>
<th>Random</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>$\text{MSB}_R/\text{MSW}$</td>
<td>$\text{MSB}_R/\text{MSI}$</td>
<td>$\text{MSB}_R/\text{MSW}$</td>
</tr>
<tr>
<td>Column</td>
<td>$\text{MSB}_C/\text{MSW}$</td>
<td>$\text{MSB}_C/\text{MSI}$</td>
<td>$\text{MSB}_C/\text{MSI}$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$\text{MSI}/\text{MSW}$</td>
<td>$\text{MSI}/\text{MSW}$</td>
<td>$\text{MSI}/\text{MSW}$</td>
</tr>
</tbody>
</table>
  – “The effect of masculine and feminine brand names on the perceived taste of a cigarette”

• Objective: assess the degree of acceptance of boys and girls into two new brands of cigarettes. Of these, the one had a name with a strong masculine connotation, "Frontiersman", and the other one with a strong feminine connotation, "April". Strong interactions showed how the brand name influenced much customers.
EXAMPLE

- Dependent variable: intention to buy (1 = not interested, 7 = sure to buy)
- Each combination of treatment was verified over a sample of 50 people. In total, 4 groups, two only male and two only female.
Factor are both significant, but the prevalent effect is the interaction (p-value almost null)

Brand= **fixed variable**
### Example

<table>
<thead>
<tr>
<th>Variance</th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
<th>$F_{\text{calc}}$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{SSB}_{R} (\text{brand})$</td>
<td>29.64</td>
<td>1</td>
<td>29.64</td>
<td>0.20</td>
<td>73%</td>
</tr>
<tr>
<td>$\text{SSB}_{C} (\text{sex})$</td>
<td>23.80</td>
<td>1</td>
<td>23.80</td>
<td>5.61</td>
<td>1.9%</td>
</tr>
<tr>
<td>$\text{SSI}_{R,C}$</td>
<td>146.24</td>
<td>1</td>
<td>146.24</td>
<td>34.48</td>
<td>0%</td>
</tr>
<tr>
<td>$\text{SSW}$</td>
<td>831.04</td>
<td>196</td>
<td>4.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1030.7</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Brand** = random variable

$\text{SSB}_{R}$ is no more significant
ABSENCE OF REPETITIONS

- Based on the statistical model, that bound the general value to the global mean, the row and column deviation from that value, the interaction and the error...

\[ Y_{ijk} = \mu + \rho_i + \tau_j + I_{ij} + \varepsilon_{ijk} \]

- ... Squaring and summing over \( i, j, k \):

\[
\sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{...})^2 = n \cdot C \cdot \sum_{i=1}^{R} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + n \cdot R \cdot \sum_{j=1}^{C} (\bar{Y}_{.j} - \bar{Y}_{...})^2 + \\
+ n \cdot \sum_{i=1}^{R} \sum_{j=1}^{C} (\bar{Y}_{i.j} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 + \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{i..})^2
\]
ABSENCE OF REPETITIONS

• TSS = SSB_R + SSB_C + SSI + SSW

\[ \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{n} (\bar{Y}_{ijk} - \bar{Y}_{ij}).^2 \]

• Without repetitions n = 1 and \( Y_{ijk} = Y_{ij} \).

• Therefore we have SSW = 0, and his DoF too,

• RC(n-1)= RC(1-1) = 0, consequently MSW, the estimator of the error usually used in the analysis is not computable (0/0)
**ABSENCE OF REPETITIONS**

- Renouncing the third index \((k)\), we have: \(Y_{ijk} = \overline{Y}_{ij}\)
- The statistical model become \((i = 1, \ldots, R; j = 1, \ldots, C)\):
  \[Y_{ij} = \mu + \rho_i + \tau_j + I_{ij}\]
- Considering the estimators of the mentioned values we have:
  \[Y_{ij} = \overline{Y}_{..} + (\overline{Y}_{i..} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (\overline{Y}_{ij} - \overline{Y}_{i..} - \overline{Y}_{.j} + \overline{Y}_{..})\]
- Moving the *Grand Mean* on the left side
  \[Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i..} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (\overline{Y}_{ij} - \overline{Y}_{i..} - \overline{Y}_{.j} + \overline{Y}_{..})\]
ABSENCE OF REPETITIONS

• Squaring and summing over i, j, we have:

\[
\sum_{i=1}^{R} \sum_{j=1}^{C} (Y_{ij} - \bar{Y}_{..})^2 = C \cdot \sum_{i=1}^{R} (\bar{Y}_{i.} - \bar{Y}_{..})^2 +
\]

\[
+ R \cdot \sum_{j=1}^{C} (\bar{Y}_{.j} - \bar{Y}_{..})^2 +
\]

\[
+ \sum_{i=1}^{R} \sum_{j=1}^{C} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2
\]

\[
\text{TSS} = \text{SSB}_{R} + \text{SSB}_{C} + \text{SSI}
\]
ABSENCE OF REPETITIONS

TSS = SSB_R + SSB_C + SSI

ABSENCE OF REPETITIONS
DoF: RC – 1 = (R – 1) + (C – 1) + (C – 1)(R – 1)
The expected values of the MSQ are:
E(MSB_R) = \sigma^2 + V_R
E(MSB_C) = \sigma^2 + V_C
E(MSI) = \sigma^2 + V_l

• No ratio would be able to isolate the effects that we want to evaluate.
• If we can assume V_l = 0, we have E(MSI) = \sigma^2.
ABSENCE OF REPETITIONS

• Under this hypothesis we can use MSI as estimator of the error instead of MSW.
• We can therefore say that, in absence of interactions, the term MSI is not null due to the experimental error. Estimation of the effects:

<table>
<thead>
<tr>
<th></th>
<th>Row</th>
<th>MSB_R/MSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
<td>MSB_C/MSI</td>
<td></td>
</tr>
</tbody>
</table>
ABSENCE OF REPETITIONS

• Usually we correct the statistical model ...

\[ Y_{ij} = \mu + \rho_i + \tau_j + \varepsilon_{ij} \]

• ... and the sum of the squares :

\[ TSS = SSB_R + SSB_C + SSW \]
ABSENCE OF REPETITIONS: EXAMPLE

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccc}
Y_{ij} & 1 & 2 & 3 \\
B & 10 & 6 & 8 \\
Y_{.j} & 4 & 6 & 6 \\
Y_{..} & 4,3 & 7 & 6 \\
Y_{i.} & 8 & 8 & 6 \\
\end{array}
\]
The value of p-value is very low, this suggests that the factor of row and column are highly significant (with a confidence level of almost 100%)
ABSENCE OF REPETITIONS: EXAMPLE

• What if the assumption of null (or negligible) interaction was wrong?
• Is the analysis solid from this point of view (gives anyway a reliable result)? Or the achieved result is completely compromised.
• We were expecting a ratio like:

\[
\frac{MSB_R}{MSW} = \frac{\sigma^2 + V_R}{\sigma^2}
\]

• But the actual result is:

\[
\frac{MSB_R}{MSW} = \frac{\sigma^2 + V_R}{\sigma^2 + V_I}, V_I > 0
\]
ABSENCE OF REPETITIONS: EXAMPLE

• Accordingly, should fail the assumption made on the interaction, it would mean that the $F_{\text{calc}}$ found is underestimated.
  – In this sense, the test is more conservative.
  – Consequently, if, despite the underestimation, it was possible to reject $H_0$, the result remains valid.

• If, on the other hand $H_0$ is accepted, some doubt may arise to be committing a type II error (say that the differences are only random, when we actually have a significance).
Questions