Framework of lesson 4:

• Project with more than two factors
• Latin Squares and analysis of variance
• Greek-Latin Squares and analysis of variance
• Practical examples
• The concept of orthonormality
• Decomposition of effects
• Practical examples
PROJECT WITH MORE THAN TWO FACTORS

• Up to now we have studied monofactorial and two factors projects. In multiple applications you may be required to assess the influence of more than two factors. In such case the number of possible combinations (number of levels raised to the number of factors) increases exponentially with the increase of the factors.

• For example if you have two factors, each with 3 levels, total combinations are $3^2$, that is 9, if the factors are 3, switching to $3^3$ (27) combinations, if you go up to four, you get even to $3^4$ (81) combinations. The number of experiments to be performed becomes prohibitive, if one includes the repetitions to be performed
PROJECT WITH MORE THAN TWO FACTORS

• Such a rise of combinations (treatment combinations) is usually a problem, because they are associated with the number of experiments to be performed, the set-up costs and test execution and data processing. Therefore, if you are forced to evaluate many factors, we resort to some techniques, which allow to limit the number of combinations to be considered in the tests.

• Latin and Greek-Latin Squares
• Project $2^k e 2^{k-p}$
• These techniques have both advantages and disadvantages, often complementary to each other.
PROJECT WITH MORE THAN TWO FACTORS

• 3 factors  →  Latin Squares
• 4 or more factors  →  Greek-Latin Squares

• Any number of level, but the same for every factor
• Example: factorial plan with three factors, each with three levels: you have $3^3$ (27) combinations, corresponding to $27 \cdot n$ experiments to be performed, $n$ is the number of repetitions. The 27 combinations can be represented as a cubical matrix of dimensions $3 \times 3 \times 3$, or as a matrix $9 \times 3$: 
It would be interesting to assess the effects of the three factors, considering not all 27 combinations, but only one third of them, 9.
PROJECT WITH MORE THAN TWO FACTORS

<table>
<thead>
<tr>
<th></th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>C₁</td>
<td>C₁</td>
<td>C₁</td>
</tr>
<tr>
<td>A₂</td>
<td>C₁</td>
<td>C₁</td>
<td>C₁</td>
</tr>
<tr>
<td>A₃</td>
<td>C₁</td>
<td>C₁</td>
<td>C₁</td>
</tr>
<tr>
<td>A₁</td>
<td>C₂</td>
<td>C₂</td>
<td>C₂</td>
</tr>
<tr>
<td>A₂</td>
<td>C₂</td>
<td>C₂</td>
<td>C₂</td>
</tr>
<tr>
<td>A₃</td>
<td>C₂</td>
<td>C₂</td>
<td>C₂</td>
</tr>
<tr>
<td>A₁</td>
<td>C₃</td>
<td>C₃</td>
<td>C₃</td>
</tr>
<tr>
<td>A₂</td>
<td>C₃</td>
<td>C₃</td>
<td>C₃</td>
</tr>
<tr>
<td>A₃</td>
<td>C₃</td>
<td>C₃</td>
<td>C₃</td>
</tr>
</tbody>
</table>

First “Floor”

Second “Floor”

Third “Floor”

Considering only the combinations in **bold**, you get the following plan 3 x 3:
LATIN SQUARES

- A: raw factor, B: column factor, C inside factor
- We have $3^2 = 9$ treatments $A_1B_1C_1$, $A_2B_1C_2$, $A_3B_1C_3$, ...
- $A_3B_3C_2$ : according to this particular plan (in fact, the Latin square), you can study three factors to the "price" of 2.
- However, the Latin square must be set up (the 9 treatment combinations must be chosen) in accordance with precise rules, which refer to a good "balance" in the distribution of treatments
LATIN SQUARES

- “Balance” rules:
- All the factors have the same number of levels (m), without limitations on them, it follows that the matrix of combinations of treatments is necessarily square.
- Each factor is assessed at each level the same number of times, m (3 in this case).
- Each level of each factor is combined with any other level of any other factor only once.
LATIN SQUARES

Meaning of “Balance”

Each line sets a level of the factor A and presents a result for each level of B and each level C. Therefore the differences between the averages of rows identify the effect of A. If also for hypotheses the third level of C (C₃) increase much the result, this contribution would be balanced, since C₃ appears on each line. And then each line would suffer from the "same" increase.
### LATIN SQUARES

<table>
<thead>
<tr>
<th></th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>C₁</td>
<td>C₁</td>
<td>C₁</td>
</tr>
<tr>
<td>A₂</td>
<td>C₂</td>
<td>C₂</td>
<td>C₂</td>
</tr>
<tr>
<td>A₃</td>
<td>C₃</td>
<td>C₃</td>
<td>C₃</td>
</tr>
</tbody>
</table>

**Meaning of “Balance”**:
Suppose you have chosen another set of combinations: C₁ combined with A₁ 3 times, the same C₂ and C₃ with A₂ and A₃. In this case, if C₃ greatly increase the response, the third row would have a much higher average of the other two. Who analyzes the data could not realize if the difference between the lines is due to the effect of A or to that of C. The effects would be confused (confounded).
Meaning of “Balance”:

All this applies assuming that there are no interactions. If there were, this would upset the mechanism for evaluating the effects. For example, suppose that the conjunction of B1 with C1 face a much higher response (positive interaction). In this case the differences between the rows may show an effect linked to A, when in fact what is observed would be only the effect of the interaction.
LATIN SQUARES

• Positive aspects related to neglect the interaction:
• The number of treatment combinations is greatly decreased. An analysis of three factors can be carried out with the same resources of an analysis to only two factors. If you have three levels, it goes from $3^3$ (27) combinations to $3^2$ (9), if there are 4, one passes from $4^3$ (64) $4^2$ (16).

• Negative aspects:
• If the hypothesis of negligible interactions was incorrect, there would be some heavy implications. In this case it would be wrong to determine the individual effects and not only the result of the F-test of significance (as seen in the case of the plan to two factors without repetitions)
LATIN SQUARES: EXAMPLE

- **Factors (4 levels):**
  - A: assistance under guarantee
  - B: opening schedule
  - C: free gadget

<table>
<thead>
<tr>
<th></th>
<th>B&lt;sub&gt;1&lt;/sub&gt;</th>
<th>B&lt;sub&gt;2&lt;/sub&gt;</th>
<th>B&lt;sub&gt;3&lt;/sub&gt;</th>
<th>B&lt;sub&gt;4&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&lt;sub&gt;1&lt;/sub&gt;</td>
<td>C&lt;sub&gt;4&lt;/sub&gt;: 855</td>
<td>C&lt;sub&gt;3&lt;/sub&gt;: 877</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;: 890</td>
<td>C&lt;sub&gt;1&lt;/sub&gt;: 997</td>
</tr>
<tr>
<td>A&lt;sub&gt;2&lt;/sub&gt;</td>
<td>C&lt;sub&gt;1&lt;/sub&gt;: 962</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;: 817</td>
<td>C&lt;sub&gt;3&lt;/sub&gt;: 845</td>
<td>C&lt;sub&gt;4&lt;/sub&gt;: 776</td>
</tr>
<tr>
<td>A&lt;sub&gt;3&lt;/sub&gt;</td>
<td>C&lt;sub&gt;3&lt;/sub&gt;: 848</td>
<td>C&lt;sub&gt;4&lt;/sub&gt;: 841</td>
<td>C&lt;sub&gt;1&lt;/sub&gt;: 784</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;: 776</td>
</tr>
<tr>
<td>A&lt;sub&gt;4&lt;/sub&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;: 831</td>
<td>C&lt;sub&gt;1&lt;/sub&gt;: 952</td>
<td>C&lt;sub&gt;4&lt;/sub&gt;: 806</td>
<td>C&lt;sub&gt;3&lt;/sub&gt;: 871</td>
</tr>
</tbody>
</table>

**Answer:**

Number of sold car
LATIN SQUARES: STATISTICAL MODEL

• Absence of repetition: $Y_{ijk} = \mu + \rho_i + \tau_j + \gamma_k + \varepsilon_{ijk}$

  Real Global Mean  Error

• $\rho_i, \tau_j, \gamma_k = \text{deviation of the true means (row, columns and inside factor) from the Grand Mean}$

• $i = 1, \ldots, m$ (row)
• $j = 1, \ldots m$ (columns)
• $k = 1, \ldots m$ (inside factor)
LATIN SQUARES: STATISTICAL MODEL

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

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

Global variability among results
Variability related to the “row effect”
Variability related to the “column effect”
Variability related to the “inside factor”

With:

\[ R = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} - \bar{Y}_{..k} + 2\bar{Y}_{..} \]
LATIN SQUARES: STATISTICAL MODEL

\[ R = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} - \bar{Y}_{.k} + 2\bar{Y}_{..} = \]

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

Difference between each cell and the grand mean

Adjustment related to row, column and inside factor effects

It is a term that has the shape of an interaction. In the absence of interactions (supposed null), it can be said that this term is non-zero only for effect of the error, the experimental uncertainty
Performing squares and summing on i, j, k, we obtain the following plan for the analysis of variance:

\[
TSS = SSB_R + SSB_C + SSB_{\text{ins.fact}} + SSW
\]

The DoF of \(SSB_R\), \(SSB_C\), \(SSB_{\text{ins.fact}}\) are equal to \((m - 1)\),

The total DoF are given by the number of all the experimental data – 1

We can obtain the DoF of the error (SSW) by difference:

\[(m^2 - 1) - 3 \cdot (m - 1) = (m - 1)(m + 1) - 3 \cdot (m - 1) = (m - 1)(m - 2)\]
LATIN SQUARES: STATISTICAL MODEL

Performing squares and summing on \(i, j, k\), we obtain the following plan for the analysis of variance:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>(m \sum_{i=1}^{m} (\bar{Y}<em>{i..} - \bar{Y}</em>{...})^2)</td>
<td>m-1</td>
<td>(\frac{SSB_R}{m - 1})</td>
</tr>
<tr>
<td>Columns</td>
<td>(m \sum_{j=1}^{m} (\bar{Y}<em>{.j} - \bar{Y}</em>{...})^2)</td>
<td>m-1</td>
<td>(\frac{SSB_C}{m - 1})</td>
</tr>
<tr>
<td>Inside factor</td>
<td>(m \sum_{i=1}^{m} (\bar{Y}<em>{.k} - \bar{Y}</em>{...})^2)</td>
<td>m-1</td>
<td>(\frac{SSB_{ins.fact.}}{m - 1})</td>
</tr>
<tr>
<td>Error</td>
<td>By subtraction</td>
<td>(m-1)(m-2)</td>
<td>(\frac{SSW}{(m-1)(m-2)})</td>
</tr>
<tr>
<td>Total</td>
<td>(\sum_{i,j,k} (\bar{Y}<em>{ijk} - \bar{Y}</em>{...})^2)</td>
<td>(m^2-1)</td>
<td></td>
</tr>
</tbody>
</table>
Performing squares and summing on $i, j, k$, we obtain the following plan for the analysis of variance:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>MSQ</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>$\frac{SSB_R}{m-1}$</td>
<td>$\sigma^2 + V_{\text{row}}$</td>
</tr>
<tr>
<td>Columns</td>
<td>$\frac{SSB_C}{m-1}$</td>
<td>$\sigma^2 + V_{\text{col}}$</td>
</tr>
<tr>
<td>Inside factor</td>
<td>$\frac{SSB_{\text{ins.fact.}}}{m-1}$</td>
<td>$\sigma^2 + V_{\text{ins.fact.}}$</td>
</tr>
<tr>
<td>Error</td>
<td>$\frac{SSW}{(m-1)(m-2)}$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
Latin Squares: Statistical Model

- Performing the square and summing over i, j, k, we obtain the following plan for the analysis of variance:
- The column of the expected values indicate clearly which ratio to calculate, to isolate the effects of row, column, and inside factors.

\[
\begin{align*}
F_{\text{calc}_{\text{row}}} &= \frac{MSB_R}{MSW} \\
F_{\text{calc}_{\text{columns}}} &= \frac{MSB_C}{MSW} \\
F_{\text{calc}_{\text{ins. fact.}}} &= \frac{MSB_{\text{ins. fact.}}}{MSW}
\end{align*}
\]
LATIN SQUARES: EXAMPLE

- FACTORS (4 levels):
  - A: assistance in guaranty
  - B: opening schedule
  - C: free gadget

ASWER:
- number of sold car
It was found that none of the three factors is significant at the threshold of $\alpha = 5\%$, however, the third factor (gadgets) has a $p$-value of about 10% (significant at the 10% threshold)
LATIN SQUARES: WITH REPETITIONS

• It's rare to perform Latin squares with repetitions (n is the number of repetitions)

• $SSB_R$, $SSB_C$, $SSB_{ins.fact.}$ are calculated the same way, with the difference that the sum of squares has to be multiplied by $(n \cdot m)$, instead of only for m. For example $SSB_R$ is expressed as:

$$SSB_R = nm \sum_{i=1}^{m} (Y_{i..} - \bar{Y}_{..})^2$$

• La SSW is always calculated by difference, knowing TSS, $SSBR$, $SSBC$, $SSB_{fatt.int.}$. By difference can also be computed the DoF of SSW: $(nm^2 – 1) – 3 \cdot (m – 1) = nm^2 – 3m + 2$
GREEK-LATIN SQUARES

• If you have more than three factors you can use Greek-Latin squares: they are called like this because once they used the letters of the Latin alphabet to indicate the third factor and those of the Greek to indicate the fourth.

• We apply the same rules already seen about Latin squares, and also the same assumption regarding the absence of interaction.

• Suppose you have four factors: A (row factor), B (column factor), C and D (inside factors). Necessary but not sufficient condition, to have a Greek-Latin square is that A, B and C form a Latin square and so did A, B and D.
GREEK-LATIN SQUARES

• In the case of Latin squares we have three factors with any number of levels. In the case of Greek-Latin squares the number of levels set the maximum number of factors that can be considered.

• If the levels are \( m \), the total DoF are \( (m^2 - 1) \). Each factor take off \( (m - 1) \) DoF. Therefore, the maximum number of the factors is given by:

\[
\frac{m^2 - 1}{m - 1} = m + 1
\]

• If a Greek-Latin square contains the maximum number of factors, it is said to be complete, otherwise it is called incomplete square
GREEK-LATIN SQUARES: EXAMPLE

- Complete Greek-Latin square: 3 levels (m = 3), 4 factors:
  - A, row factor, B, column factor and C and D, inside factors

- Strong reduction of the number of combinations: from $3^4$ (81) to $3^2$ (9: $A_1B_1C_1D_1$, $A_1B_2C_2D_2$, ..., $A_3B_3C_2D_1$): a 4 factor plan become a 2 factor plan.
GREEK-LATIN SQUARES: EXAMPLE

- Another complete Greek-Latin square: 5 levels (m = 5), 6 factors:
  - A, row factor, B, column factor and C, D, E, F, inside factors

<table>
<thead>
<tr>
<th></th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
<th>B₄</th>
<th>B₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>C₁D₁E₁F₁</td>
<td>C₂D₂E₂F₂</td>
<td>C₃D₃E₃F₃</td>
<td>C₄D₄E₄F₄</td>
<td>C₅D₅E₅F₅</td>
</tr>
<tr>
<td>A₂</td>
<td>C₂D₃E₄F₅</td>
<td>C₃D₄E₅F₁</td>
<td>C₄D₅E₁F₂</td>
<td>C₅D₁E₂F₃</td>
<td>C₁D₂E₃F₄</td>
</tr>
<tr>
<td>A₃</td>
<td>C₃D₅E₂F₄</td>
<td>C₄D₁E₃F₅</td>
<td>C₅D₂E₄F₁</td>
<td>C₁D₃E₅F₂</td>
<td>C₂D₄E₁F₃</td>
</tr>
<tr>
<td>A₄</td>
<td>C₄D₂E₅F₃</td>
<td>C₅D₃E₁F₄</td>
<td>C₁D₄E₂F₅</td>
<td>C₂D₅E₃F₁</td>
<td>C₃D₁E₄F₂</td>
</tr>
<tr>
<td>A₅</td>
<td>C₅D₄E₃F₂</td>
<td>C₁D₅E₄F₃</td>
<td>C₂D₁E₅F₄</td>
<td>C₃D₂E₁F₅</td>
<td>C₄D₃E₂F₁</td>
</tr>
</tbody>
</table>

- Strong reductions of the number of combinations: from $5^6 \ (15625)$ to $5^2 \ (25: \ A_1B_1C_1D_1E_1F_1, \ ... , \ A_5B_5C_4D_3E_2F_1)$
GREEK-LATIN SQUARES

• Complete Greek-Latin squares hide a problem. In absence of repetitions, all the DoF made available from the set of experimental data are used for the effects related to the factors. None remains for the error.

• In this case the analysis of variance cannot be performed (MSW = 0/0). For instance, in the previous case we have:
  • Total DoF (TSS) = 5^2 − 1 = 24
  • DoF of each factor = 5 − 1 = 4
  • Total 6 factors (A, B, C, D, E, F): 24 − 6·4 = 24 − 24 = 0
  • To work around this problem, you can use repetition, but more often it is preferable to use incomplete squares, where, in front of m levels, we consider m factors or less.
GREEK-LATIN SQUARES: EXAMPLE

• Incomplete Greek-Latin square: 5 levels (m = 5), 4 factors:
• A, row factor, B, column factor and C and D inside factors

<table>
<thead>
<tr>
<th></th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
<th>B₄</th>
<th>B₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>C₁D₁</td>
<td>C₂D₂</td>
<td>C₃D₃</td>
<td>C₄D₄</td>
<td>C₅D₅</td>
</tr>
<tr>
<td>A₂</td>
<td>C₂D₃</td>
<td>C₃D₄</td>
<td>C₄D₅</td>
<td>C₅₁</td>
<td>C₁₂</td>
</tr>
<tr>
<td>A₃</td>
<td>C₃D₅</td>
<td>C₁D₁</td>
<td>C₅D₂</td>
<td>C₁D₃</td>
<td>C₂D₄</td>
</tr>
<tr>
<td>A₄</td>
<td>C₄D₂</td>
<td>C₅D₃</td>
<td>C₁D₄</td>
<td>C₂D₅</td>
<td>C₃D₁</td>
</tr>
<tr>
<td>A₅</td>
<td>C₅D₄</td>
<td>C₁D₅</td>
<td>C₂D₁</td>
<td>C₃D₂</td>
<td>C₄D₃</td>
</tr>
</tbody>
</table>

• Strong reductions of the number of combinations: from $5^4$ (625) to $5^2$ (25: $A₁B₁C₁D₁$, $A₁B₂C₂D₂$, ..., $A₅B₅C₄D₃$)
GREEK-LATIN SQUARES: EXAMPLE

• ANOVA plan in simbolic form:

<table>
<thead>
<tr>
<th></th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
<th>$F_{\text{calc}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>$SSB_A$</td>
<td>4</td>
<td>$SSB_A/4$</td>
<td>$MSB_A/MSW$</td>
</tr>
<tr>
<td>Factor B</td>
<td>$SSB_B$</td>
<td>4</td>
<td>$SSB_B/4$</td>
<td>$MSB_B/MSW$</td>
</tr>
<tr>
<td>Factor C</td>
<td>$SSB_C$</td>
<td>4</td>
<td>$SSB_C/4$</td>
<td>$MSB_C/MSW$</td>
</tr>
<tr>
<td>Factor D</td>
<td>$SSB_D$</td>
<td>4</td>
<td>$SSB_D/4$</td>
<td>$MSB_D/MSW$</td>
</tr>
<tr>
<td>Error</td>
<td>SSW</td>
<td>8</td>
<td>SSW/8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>TSS</td>
<td>24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Obtained by subtraction
ORTHOGONALITY

• Stepping back, we studied the single-factor analysis of variance and showed how one can investigate the differences between the columns.
• It was shown how the tests two by two may lead to inconsistencies (no transitivity) and how it can be heavily increased the probability of error.
• Another method to investigate the effect of a factor is to decompose the $SSB_C$ in $(C-1)$ component, as many as the DoF.
ORTHOGONALITY

• Each of the \((C - 1)\) contributions corresponds to a question, to which give an answer by a statistical test. As will be seen, the formulation of a question is equivalent to writing of linear combinations between the averages of the scores.

• If any "question" is "orthogonal" to the other, then apply the following properties:
  – Each question is independent from the others and it is interesting how the chances of making errors of type I or II during each test are independent of each other
  – The \((C - 1)\) contributions, also of SSQ, are a decomposition of \(SSB_C\): their sum equal it perfectly
• As already described, each $SSB_C$ can be decomposed in $(C - 1)$ SSQ, as many as the DoF. Each SSQ corresponds to a linear combination of column means, in turn linked to a question.

• So, given the questions, on the basis of these, we can construct with suitable criteria the linear combinations, and from these can be calculated SSQ, then verifying that their sum equals the $SSB_C$. Assuming that there are 4 columns (4 levels), the components of $SSB_C$ are 3, as well as linear combinations that can be written.
LINEAR COMBINATIONS

• For simplicity we assume: \( \overline{Y}.j = Y_j \)

• Linear combination are of the following kind:

\[
Z_1 = a_{11}Y_1 + a_{12}Y_2 + a_{13}Y_3 + a_{14}Y_4 = \sum_{j=1}^{4} a_{1j}Y_j
\]

\[
Z_2 = a_{21}Y_1 + a_{22}Y_2 + a_{23}Y_3 + a_{24}Y_4 = \sum_{j=1}^{4} a_{2j}Y_j
\]

\[
Z_3 = a_{31}Y_1 + a_{32}Y_2 + a_{33}Y_3 + a_{34}Y_4 = \sum_{j=1}^{4} a_{3j}Y_j
\]
LINEAR COMBINATIONS

• So we have the general form: 
  \[ Z_i = \sum_{j=1}^{4} a_{ij} Y_j \]

• In matrix form we have as following: the matrix of the coefficients typically have as many rows as are the DoF (C-1) and as many columns as are the levels (C)

\[
\begin{bmatrix}
Z_1 \\
Z_2 \\
Z_3
\end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24} \\
a_{31} & a_{32} & a_{33} & a_{34}
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
Y_4
\end{bmatrix}
\]
LINEAR COMBINATIONS

• In order for the decomposition of $SSB_C$ to be exact, it is necessary that:

$$R \cdot \sum_{i=1}^{3} Z_i^2 = R \cdot \sum_{i=1}^{3} (Y_j - \bar{Y..})^2 \iff \sum_{i=1}^{3} Z_i^2 = \sum_{i=1}^{3} (Y_j - \bar{Y..})^2$$

• For this to occur, we must fulfill three rules:
  • Orthogonality
  • Orthonormality
  • “balanced” question, from the mathematical point of view
LINEAR COMBINATIONS

• Orthogonality:

• The scalar product between each row vector of the matrix of the coefficients is zero. For example, whereas the first two lines:

\[ <a_{1j}, a_{2j}> = 0 \iff \]

\[ a_{11} \cdot a_{21} + a_{12} \cdot a_{22} + a_{13} \cdot a_{23} + a_{14} \cdot a_{24} = \sum_{j=1}^{4} a_{1j} \cdot a_{2j} = 0 \]

• In general:

\[ \sum_{j=1}^{4} a_{1j} \cdot a_{2j} = 0, \quad \forall i_1, i_2, \quad i_1 \neq i_2 \]
LINEAR COMBINATIONS

• Orthonormality:
• Each row vector of the matrix of coefficients should have unitary norm. The norm of a vector is typically given by the root square of the scalar product of the vector itself. For the vector of the i-th row, you have:

\[
\|a_{ij}\| = \sqrt{<a_{ij},a_{ij}>} = \sqrt{\sum_{j=1}^{4} a_{ij}^2}
\]

• General rule:

\[
\sum_{j=1}^{4} a_{ij}^2 = 1
\]
LINEAR COMBINATIONS

• Balance from the mathematical point of view:
  • The sum of the coefficients (non-square) of each line must be zero. We will see from the examples what this implies in terms of mathematical balance. For the i-th row:

\[
\sum_{j=1}^{4} a_{ij} = 0
\]

• Of these three rules, the only essential are the first and the third. The requirement orthonormality is essential, but deserves minor concerns. If a vector is orthogonal to the other and the sum of its elements is zero, it is sufficient to divide by the norm, to normalize it.
ORTHOGONALITY: EXAMPLE 1

• This is an example of orthonormal matrix:

\[
\begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}
\end{bmatrix}
\]

• Verify that meets the three properties:

• Orthogonality:

\[
\langle a_{1j}, a_{2j} \rangle = \frac{1}{4} - \frac{1}{4} - \frac{1}{4} + \frac{1}{4} = 0
\]
\[
\langle a_{1j}, a_{3j} \rangle = \frac{1}{4} - \frac{1}{4} + \frac{1}{4} - \frac{1}{4} = 0
\]
\[
\langle a_{2j}, a_{3j} \rangle = \frac{1}{4} + \frac{1}{4} - \frac{1}{4} - \frac{1}{4} = 0
\]
ORTHOGONALITY: EXAMPLE 1

• Orthonormality: \[ |a_{1j}|^2 = |a_{2j}|^2 = |a_{3j}|^2 = 4 \cdot \frac{1}{4} = 1 \]

• Balance:

\[
\begin{align*}
\sum_{j=1}^{4} a_{1j} &= \frac{1}{2} + \frac{1}{2} - \frac{1}{2} - \frac{1}{2} \\
\sum_{j=1}^{4} a_{2j} &= \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} \\
\sum_{j=1}^{4} a_{3j} &= \frac{1}{2} - \frac{1}{2} - \frac{1}{2} + \frac{1}{2}
\end{align*}
\]

• Let’s suppose to have the following column averages (4 repetitions, R = 4):

<table>
<thead>
<tr>
<th>( Y_1 )</th>
<th>( Y_2 )</th>
<th>( Y_3 )</th>
<th>( Y_4 )</th>
<th>( Y_\cdot )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>-3</td>
<td>2</td>
</tr>
</tbody>
</table>
ORTHOGONALITY: EXAMPLE 1

\[
SSB_C = R \cdot \sum_{j=1}^{C=4} (Y_j - \overline{Y..})^2 = \\
= 4 \cdot \left[ (6 - 2)^2 + (4 - 2)^2 + (1 - 2)^2 + (-3 - 2)^2 \right] = 46 \cdot R = 184
\]

<table>
<thead>
<tr>
<th></th>
<th>(Y_1)</th>
<th>(Y_2)</th>
<th>(Y_3)</th>
<th>(Y_4)</th>
<th>(Z)</th>
<th>(Z^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z_1)</td>
<td>1/2</td>
<td>1/2</td>
<td>-1/2</td>
<td>-1/2</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>(Z_2)</td>
<td>1/2</td>
<td>-1/2</td>
<td>1/2</td>
<td>-1/2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>(Z_3)</td>
<td>1/2</td>
<td>-1/2</td>
<td>-1/2</td>
<td>1/2</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>
ORTHOGONALITY: EXAMPLE 1

• Therefore, the three decompositions are:

\[ SSQ_1 = Z_1^2 \cdot R = 36 \cdot R = 144 \quad SSQ_2 = Z_2^2 \cdot R = 9 \cdot R = 36 \]
\[ SSQ_3 = Z_3^2 \cdot R = 1 \cdot R = 4 \]

• Their sum is equal to \( SSB_C \) and such a result is related to the choice of the right matrix that respect the three required properties:

\[
\sum_{i=1}^{3} SSQ_i = R \sum_{i=1}^{3} Z_i^2 = R \cdot (36 + 9 + 1) = 46R = 184
\]
It was performed a statistical analysis to assess the impact of different drugs on the relief of people suffering from headaches. Has been performed a monofactorial ANOVA with: X= medicine (4 different types, i.e. 4 levels); Y= reduced pain (1 = no effect = 15 disappearance of the disorder). 8 people were administered the 4 medicines, then R = 8.

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Aspirin 1</th>
<th>Aspirin 2</th>
<th>Orudis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>Y₁=5</td>
<td>Y₂=6</td>
<td>Y₃=7</td>
<td>Y₄=10</td>
</tr>
</tbody>
</table>
• We reject $H_0$: drugs work. We want now to investigate which of them should be held responsible for the differences found (not all $\mu_j$ are equal).

• The questions are:

1. Is placebo actually different from the other drugs together? (P vs. P’)
2. Does exist any difference between the two Aspirines? (A1 vs. A2)
3. Does exist any difference between Orudis and the two Aspirins together)? (A vs. O)

<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>Medicine</td>
<td>112</td>
<td>3</td>
<td>37.33</td>
<td>7.47</td>
<td>0.0008</td>
</tr>
<tr>
<td>Error</td>
<td>140</td>
<td>28</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>252</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ORTHOGONALITY: EXAMPLE 2

- Questions ↔ Linear combinations among means of column $Y_j$
- To compare Placebo with the other three drugs together is the same as subtract from the mean of placebo the means of the other three.
- To compare two aspirins means subtract the mean of one from the mean of the other
- Working in this way we obtain the following matrix of coefficients

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>A1</th>
<th>A2</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>P vs. P'</td>
<td>-1</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>A1 vs. A2</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A vs. O</td>
<td>0</td>
<td>-1/2</td>
<td>-1/2</td>
<td>1</td>
</tr>
</tbody>
</table>
ORTHOGONALITY: EXAMPLE 2

• Note that there was no point in subtracting the placebo from the average of the other three. In that way, the coefficients would be (1 -1 -1 -1) and failure to comply with the rule of balance would signal an anomaly.
• They are therefore respected the rule of orthogonality and the balance. It is not respected the rule of orthonormality: none of the three row vectors has unitary module. To normalize, simply divide each to its norm.

\[ \|a_{ij}\| = \sqrt{\langle a_{ij}, a_{ij} \rangle} = \sqrt{\sum_{j=1}^{4} a_{ij}^2} \]
ORTHOGONALITY: EXAMPLE 2

- Norm of the three row vector:

\[
\|a_{1j}\| = \sqrt{\sum_{j=1}^{4} a_{1j}^2} = \sqrt{1+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}} = \sqrt{\frac{12}{9}} = 1.15
\]

\[
\|a_{2j}\| = \sqrt{\sum_{j=1}^{4} a_{2j}^2} = \sqrt{1+1} = \sqrt{2} = 1.41
\]

\[
\|a_{3j}\| = \sqrt{\sum_{j=1}^{4} a_{3j}^2} = \sqrt{\frac{1}{4}+\frac{1}{4}+1} = \sqrt{\frac{3}{2}} = 1.22
\]

- Orthonormal matrix:

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>A1</th>
<th>A2</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>P vs. P’</td>
<td>-0.87</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>A1 vs. A2</td>
<td>0</td>
<td>-0.71</td>
<td>0.71</td>
<td>0</td>
</tr>
<tr>
<td>A vs. O</td>
<td>0</td>
<td>-0.41</td>
<td>-0.41</td>
<td>0.81</td>
</tr>
</tbody>
</table>
### ORTHOGONALITY: EXAMPLE 2

<table>
<thead>
<tr>
<th></th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
<th>$Z$</th>
<th>$Z^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_1$</td>
<td>-0.87</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>2.31</td>
<td>5.33</td>
</tr>
<tr>
<td>$Z_2$</td>
<td>0</td>
<td>-0.71</td>
<td>0.71</td>
<td>0</td>
<td>0.71</td>
<td>0.5</td>
</tr>
<tr>
<td>$Z_3$</td>
<td>0</td>
<td>-0.41</td>
<td>-0.41</td>
<td>0.81</td>
<td>2.86</td>
<td>8.17</td>
</tr>
</tbody>
</table>

$$SSQ_1 = Z_1^2 \cdot R = 5.33 \cdot 8 = 42.64$$
$$SSQ_2 = Z_2^2 \cdot R = 0.50 \cdot 8 = 4.00$$
$$SSQ_3 = Z_3^2 \cdot R = 8.17 \cdot 8 = 65.36$$

$$R \cdot \sum_{i=1}^{3} Z_i^2 = \sum_{i=1}^{3} SSQ_i = 8 \cdot (5.33 + 0.50 + 8.17) = 112 = SSB_C$$
ORTHOGONALITY: EXAMPLE 2

- Analysis of Variance:

<table>
<thead>
<tr>
<th>Medicine</th>
<th>SSQ</th>
<th>DoF</th>
<th>MSQ</th>
<th>$F_{\text{calc}}$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P vs. P'</td>
<td>42.64</td>
<td>3</td>
<td>42.64</td>
<td>8.53</td>
<td>0.007</td>
</tr>
<tr>
<td>A1 vs. A2</td>
<td>4.00</td>
<td>3</td>
<td>4.00</td>
<td>0.80</td>
<td>0.38</td>
</tr>
<tr>
<td>A vs. O</td>
<td>65.36</td>
<td>3</td>
<td>65.36</td>
<td>13.07</td>
<td>0.001</td>
</tr>
<tr>
<td>Error</td>
<td>140</td>
<td>28</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>252</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The first and in the third case are significant (p.value <5%). This means that the placebo is different from other drugs and that the two aspirins are different from Orudis. No significant differences are observed in the second case: the two aspirins are equivalent.
ORTHOGONALITY: EXAMPLE 2

• Repartition of the variance associated with the columns (the types of medicine):

• The biggest portion is related to the differences between the Orudis and aspirin, another good part is related to differences compared to placebo and, although not significant, there is also a small slice related to differences between the two aspirins.