Use of factorial design with significant interaction and Tukey's Test in agricultural and environmental experiments

Uso de planejamento fatorial com interação significativa e teste de Tukey em experimentos agrícolas e ambientais

José Tadao Yamada¹
Rafaela Simões Espírito Viana²
Ramona Rodrigues Santos³
Cleginaldo Pereira de Carvalho⁴
Maria da Glória Diniz de Almeida⁵
José Glênio Medeiros de Barros⁶
Nilo Antonio de Souza Sampaio⁷

Abstract
This study aimed to present one of the applications of complete factorial planning with significant interactions as well as Tukey's Test in the agricultural and environmental area.

¹Graduated in Metallurgical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: jtdyamada@gmail.com
Orcid: https://orcid.org/0009-0000-6481-4642
²Graduate Student in Chemical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: rafaelaviana2679@gmail.com
Orcid: https://orcid.org/0009-0008-9977-479X
³Master in Metallurgical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: ramonarodrigues@id.uff.br
Orcid: https://orcid.org/0009-0009-9389-0519
⁴Doctor in Mechanical Engineering, Universidade Estadual Paulista Júlio de Mesquita Filho, Av. Dr. Ariberto Pereira da Cunha, 333, Pedregulho, Guaratinguetá - SP, CEP: 12516-410. E-mail: cleginaldo.carvalho@unesp.br Orcid: https://orcid.org/0000-0001-7364-2096
⁵Doctor in Mechanical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: gloria_uerj@yahoo.com.br
Orcid: https://orcid.org/0000-0002-0645-8907
⁶Doctor in Mechanical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: glenio.barros@gmail.com
Orcid: https://orcid.org/0000-0002-6902-599X
⁷Doctor in Mechanical Engineering, Universidade do Estado do Rio de Janeiro, Rod. Pres. Dutra, km 298, Polo Industrial, Resende - RJ, CEP: 27537-000. E-mail: nilo.samp@terra.com.br Orcid: https://orcid.org/0000-0002-6168-785X
achieve this objective, applied research of exploratory nature and quantitative approach was carried out. As method and technical procedures were adopted, respectively, the bibliographical research, the literature review, and a case study. As a contribution of this research, it is pointed out the use of this tool in other research fields, but mainly in the environmental area.

**Keywords:** Statistical Software. Complete Factorial Planning. Tukey's Test.

**Resumo**
Este estudo teve como objetivo apresentar uma das aplicações do planejamento fatorial completo com interações significativas e do Teste de Tukey na área agrícola e ambiental. Para atingir esse objetivo, foi realizada uma pesquisa aplicada de natureza exploratória e abordagem quantitativa. Como método e procedimentos técnicos foram adotados, respectivamente, a pesquisa bibliográfica, a revisão da literatura e um estudo de caso. Como contribuição desta pesquisa, aponta-se a utilização desta ferramenta em outros campos de pesquisa, mas principalmente na área ambiental.

**Palavras-chave:** Software Estatístico. Planejamento Fatorial Completo. Teste de Tukey.

**Introduction**

Statistics today has made a significant contribution to the decision-making process, as much of what is done is based on quantitative methods, and statistics is one such area. In the information and knowledge age, statistics uses mathematics to support professionals in business, government, and researchers. Wherever there is uncertainty, this tool can be used. Statistics consists of the planning, collection, consistency, tabulation, analysis, and interpretation of data from surveys involving censuses or sample surveys (Ignácio, 2010; Rezende et al., 2023; Sampaio et al., 2010). All processes that have a potential impact on product quality must first be validated. Validation consists in proving the constancy of the results obtained under the control of certain input conditions. To ensure the robustness of the analyzed processes, validation must in turn follow systematically determined processes (Gomes et al., 2022; Liang et al., 2017; Mazza et al., 2023; Sales et al., 2022; Silva & Sant’Anna, 2007).

Currently, statistics has contributed significantly to the decision-making process because much of what is produced is based on quantitative methods, statistics being one of them. DOE is a structured and organized method used to relate different input and output...
factors of a process, involving the definition of the set of experiments, in which all relevant factors are systematically varied. The methodology of design of experiments was first applied in agricultural experiments, and quickly became used in the fields of engineering, biology, chemistry, and environment. In recent years, there is a constant concern among researchers to statistically prove the validity of laboratory experiments (Cardoso, Reis, Silva, Barros, et al., 2023; Dal Molin et al., 2008; Fonseca et al., 2023).

Statistical and mathematical techniques are useful for the development, improvement, and optimization of production processes. In this context, it is important to apply these techniques to the identification of factors that influence quality characteristics of processes or products (Vieira & Epprecht, 2009).

Experimental planning is a powerful tool to study the joint effect of several factors on a response variable of interest. One of the most used techniques is factorial design, in which k factors are involved, each of them present at different levels. The simplest case is that in which each factor k is present in only two levels. In full factorial design, each possible combination of factor levels should be tested to determine how much the experiment under study is affected by each variable. The number of experiments increases with the number of variables, so this method is not practical when more than four variables are involved (Neves et al., 2002; Rezende et al., 2023; Talati et al., 2017).

**Theoretical Referential**

When launching a new product and/or process, it is usually a matter of working with many variables. The conscious planning of the experiments that must be used to manipulate these variables and reach the desired answers, is indispensable so that reliable results can be obtained, and consistent statistical analyses can be performed. In this context, it is no longer possible to develop products and processes through trial-and-error procedures, as was done at the beginning of the last century. The strong competitiveness, the diffusion of technological processes, as well as the responsibility of the scientific community today make such procedures unfeasible. The optimization of processes and products requires more than ever a robust statistical study (Cardoso, Reis, Silva, Almeida, et al., 2023; Costa, 2011).

In general, DoE is defined as a combination of experiments (treatments) planned to enable the relationship of the effect of a set of levels of independent factors on one or more dependent response variables, judged to be of interest to the process. From these experiments, it is possible to apply statistical tests on the significance of the effects of factors, as well as develop empirical mathematical models that allow, for the considered experimental range, to
predict the effects of certain combinations of these factors on the response variables of the system. A practical way to design experiments is to construct $2^k$ factorial designs (two level experiments with $k$ factors). In cases where there are a large number of factors, it is convenient to split the experiment through the confounding technique, where main effects are mixed with the effects of higher order interactions (for example, triple interactions) to reduce costs and time of conducting the study. (Montgomery, 1991; Oprime et al., 2017).

Factorial planning is a useful analytical strategy, and its application lies in the screening of the most relevant variables of a given analytical system. After this process of screening of the most significant variables, experiments are performed to refine and allow refinement and a better understanding of the system under study. In the organization of an experimental design, it is essential that it can provide exactly the kind of information that the analyst needs. Thus, the most important activity is not exactly in the data analysis, but in the design of the experiments in which these data are to be obtained. To plan experiments is to define a sequence of experimental data collection to achieve certain objectives. Among the experimental planning methods available in the literature, factorial literature, factorial design is the most suitable when you want to study the effects of two or more influence variables, where in each trial or replicate, all possible combinations of the levels of each variable possible combinations of the levels of each variable are investigated (Chodkowski et al., 1973; Vicentini et al., 2011).

**Research Method**

This paper can be classified as an applied research, as it aims to provide improvements in the current literature, with normative empirical objectives, aiming to develop policies and strategies that improve the current condition (Espuny et al., 2022; Reis et al., 2022; Stüpp et al., 2015). The problem approach is quantitative, as is the modeling and simulation research method. The research steps were performed following the sequence shown in Figure 1.

- **Step 1:** The Experimental data was selected from the work of (Stüpp et al., 2015). This choice was since this work brings interesting data to be treated with the respective statistical analyses.
- **Step 2:** Conduct the Full Factorial Experiment with Statistica 1.0 Software.
- **Step 3:** Unfolding the Experiment and Subsequent Tukey's Test.
- **Step 4:** The interpretation of the results was performed.
- **Step 5:** The problem situation was analyzed, and a decision could be made based on the results obtained.

Revista Gestão e Secretariado (GeSec), São Paulo, SP, v. 14, n. 6, 2023, p. 9815-9828.
• Step 6: The conclusions presented were drawn from the results obtained in the previous steps.

**Figure1. Search method steps**

3.1 Case Study

The problem described by (Stüpp et al., 2015) shows the growth of a seedling of Mimosa scabrella Benth, and studies the height of the tree as a function of the type of container and the species. In the original work, this study was performed using polynomial regression. In this study, a design of experiments was used, and the morphological analysis was performed 80 days after sowing, through measurements of the height of the aerial part (cm), as a function of these two factors. There are three types of containers C1, C2 and C3 and two types of species S1 and S2. The results are presented in Table 1.

<table>
<thead>
<tr>
<th>Container</th>
<th>Species</th>
<th>Replay</th>
<th>Height (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>S1</td>
<td>1</td>
<td>26,2</td>
</tr>
<tr>
<td>C1</td>
<td>S1</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>C1</td>
<td>S1</td>
<td>3</td>
<td>25,8</td>
</tr>
<tr>
<td>C1</td>
<td>S1</td>
<td>4</td>
<td>25,4</td>
</tr>
<tr>
<td>C1</td>
<td>S2</td>
<td>1</td>
<td>24,8</td>
</tr>
<tr>
<td>C1</td>
<td>S2</td>
<td>2</td>
<td>24,6</td>
</tr>
<tr>
<td>C1</td>
<td>S2</td>
<td>3</td>
<td>26,7</td>
</tr>
<tr>
<td>C1</td>
<td>S2</td>
<td>4</td>
<td>25,2</td>
</tr>
<tr>
<td>C2</td>
<td>S1</td>
<td>1</td>
<td>25,7</td>
</tr>
<tr>
<td>C2</td>
<td>S1</td>
<td>2</td>
<td>26,3</td>
</tr>
<tr>
<td>C2</td>
<td>S1</td>
<td>3</td>
<td>25,1</td>
</tr>
</tbody>
</table>
Use of factorial design with significant interaction and Tukey's Test in agricultural and environmental experiments

<table>
<thead>
<tr>
<th>C2</th>
<th>S1</th>
<th>4</th>
<th>26.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>S2</td>
<td>1</td>
<td>19.6</td>
</tr>
<tr>
<td>C2</td>
<td>S2</td>
<td>2</td>
<td>21.1</td>
</tr>
<tr>
<td>C2</td>
<td>S2</td>
<td>3</td>
<td>19.1</td>
</tr>
<tr>
<td>C2</td>
<td>S2</td>
<td>4</td>
<td>18.6</td>
</tr>
<tr>
<td>C3</td>
<td>S1</td>
<td>1</td>
<td>22.8</td>
</tr>
<tr>
<td>C3</td>
<td>S1</td>
<td>2</td>
<td>19.4</td>
</tr>
<tr>
<td>C3</td>
<td>S1</td>
<td>3</td>
<td>18.8</td>
</tr>
<tr>
<td>C3</td>
<td>S1</td>
<td>4</td>
<td>19.2</td>
</tr>
<tr>
<td>C3</td>
<td>S2</td>
<td>1</td>
<td>19.8</td>
</tr>
<tr>
<td>C3</td>
<td>S2</td>
<td>2</td>
<td>21.4</td>
</tr>
<tr>
<td>C3</td>
<td>S2</td>
<td>3</td>
<td>22.8</td>
</tr>
<tr>
<td>C3</td>
<td>S2</td>
<td>4</td>
<td>21.3</td>
</tr>
</tbody>
</table>

Table 1. Height of seedlings as a function of container type and species.

The statistical assumptions of normality and homoscedasticity were performed by the Liliefors and Bartlett tests and a factorial experiment with significant interaction was performed to verify the interaction between the factors and after each one, the corresponding Tukey's Test to verify the best factor.

Results and Discussions

From the analysis in Table 1, it is first necessary to perform a Full Factorial Experiment for the two factors and the interaction between them. Both experiments were done using the software Statistica 1.0. The results show that both the container and the species and their interactions are significant (p-value < 0.05), with a mean square of the residual of 1.25 and a degree of freedom of 18 with 95% confidence, which will be used until the end of the analysis.

Figure 2. Factorial Experiment performed with Statistica 10.0 software
The result shows that the interaction is significant, so it does not make sense to perform the Tukey test for container and species separately. Therefore, the experiment was split, i.e., I will select the species and see which is the best container within each species and then I will select the containers and see which is the best species within each container. Initially, with Species 1 fixed, anova is performed with the height as a function of the container, which can be seen in Figure 3. The interpretation is that as the p value was less than 0.05 (5%) it was significant, which means that at least one of the containers differed from the others.

![Figure 3. Experimental Plot with Fixed S1 Species performed with Statistica 10.0 software](image1)

Next the Tukey test is performed, in which the best height averages occur with container C1 and C2 being almost equal, and container C3 being a little worse. Seen in Figure 4.

![Figure 4. Tukey's test to verify the best container with S1 Fixed Species performed with Statistica 10.0 software](image2)
The Anova with the height as a function of the container is then performed with Species 2, which can be seen in Figure 5. The interpretation is that as the p-value is less than 0.05 it is significant, which means that at least one of the containers differs from the others.

![Figure 5 Experimental Plot with Fixed S1 Species performed with Statistica 10.0 software](image)

Next the Tukey’s test is performed, in which the average height of container C1 is the best and C2 and C3 are almost equal.

![Figure 6. Tukey's test to verify the best container with S2 Fixed Species performed with Statistica 10.0 software](image)
The experiment was then further deployed, fixing container 1 as a function of the two species to verify which of them is the best for this container, which can be seen in Figure 7.

Since the p-value was 0.3383 (33.83%), greater than 0.05 (5%), it means that it was not significant and that therefore for container C1 the two species show the same or very close results, which can be verified by Tukey's Test in Figure 8.
Then, the experiment was further deployed, fixing container 2 as a function of the two species to verify which of them is the best for this container, which can be seen in Figure 9.

![Figure 9 Experimental Plot with Fixed C2 Species performed with Statistica 10.0 software](image)

The interpretation is that since the p-value is less than 0.05 it is significant, which means that within container 2 at least one of the species differs from the others. The Tukey test was then performed, as shown in Figure 9. Next, the Tukey test is performed, in which the average height of Species S1 is better than S2 seen in Figure 10.

![Figure 10 Tukey's test to verify the best container with C2 Fixed Species performed with Statistica 10.0 software](image)
Finally, we repeated the procedure by setting the C3 container to the two species to check which one is best for this container, which can be seen in Figure 11.

Since the p-value was 0.2943 (29.43%), greater than 0.05 (5%) it means that it was not significant and that therefore for the C3 container the two species show the same or very close results which can be verified by Tukey's Test in Figure 12.
Conclusion

The objective of this study was to show the importance of the tools Complete Factor Analysis with Significant Interaction and Tukey's Test to help as a tool for decision making in experiments related to agriculture and environment. To achieve this goal, several statistical treatments were performed in a case study. As a contribution of this research, it points out opportunities to use these tools in situations linked to the environmental, food, and agricultural areas. The second objective was to show the importance of Factorial Planning in research, because with the evolution of artificial intelligence, these statistical tools will contribute significantly to relevant conclusions.

References


Use of factorial design with significant interaction and Tukey's Test in agricultural and environmental experiments


Submetido em: 09.05.2023
Aceito em: 14.06.2023