Boletim Informativo Geum
Informative Geum Bulletin

Boletim Informativo Geum, v.7, n.2, p. 39-47, abr./jun.,2016
ISSN 2237-7387 (Impresso)

DEVELOPMENT OF A PRACTICAL TOOL FOR ANALYSIS OF DATA OBTAINED BY FACTORIAL DESIGN 2^2 E 2^3 USING THE SOFTWARE MICROSOFT OFFICE EXCEL®2007 (MICROSOFT CORPORATION, EUA)

Handerson Rodrigues Silva Lima[1], Laisa Lis Fontinele de Sá[2], Francisco Daniel Leal Sousa[3], Lívio César Cunha Nunes[2*]

[1]Universidade Federal do Piauí, Campus Ministro Petrônio Portella, Programa de Pós-graduação em Ciências Farmacêuticas, Teresina, Piauí, Brasil.
[2]Universidade Federal do Piauí, Campus Ministro Petrônio Portella, Programa de Pós-graduação em Biotecnologia - RENORBIO, Polo Teresina, Piauí, Brasil.
[3]Centro Universitáio Uninovafapi, Mestrado Profissional em Saúde da Família, Teresina, Piauí, Brasil.

* Correspondence: Adress:
E-mail: liviocesar@hotmail.com

ABSTRACT

The factorial design is an important statistical tool, which due to its simplicity is being increasingly used for different samples and purposes. This increasing use is justified by the allowance of results' interpretation considering all experimental parameters involved, in addition to provide the effects of possible interactions between the selected variables. The mathematical expressions used for the required calculations in the evaluation of the data obtained are widely presented in the literature and this study aimed to produce a practical tool that can assist the execution of these calculations in a fast and simple way. Spread sheets were produced for the analysis of the obtained data by factorial design 2^2 and 2^3 using the software MICROSOFT OFFICE EXCEL®2007 (MICROSOFT CORPORATION, USA). With this tool, it is possible, from the data obtained through factorial design assays, to quickly obtain the values of the each factor's effects on the response being studied, as well as their interactions, also assessing the statistical significance of each one about the response in question. It is noteworthy that the direct use of 'friendly' softwares without first knowing the methodology's fundaments, can lead users to dangerous misinterpretations. Therefore, it is recommended to previously understand the fundamentals of the methodology and the importance of performing each step in the procedures' implementation.

Keywords: factorial design, practical tool, effects, statistical analysis.

UNDERSTANDING THE FACTORIAL DESIGN

The factorial design is an important statistical tool that due to its simplicity, is being increasingly used. This increase is also occurring because this tool allows the interpretation of results considering all experimental parameters involved, making it possible to conduct a multivariate system analysis, aiming to improve all variables that compose the experimental system (COSTA; KORN; NOGUEIRA, 2006).

Thus, the factorial design allows simultaneous variables evaluation, with a reduced number of experiments, without harming the quality of the information (PEREZ et al., 2015).

Some advantages of using Factorial Design are cited by COSTA; ALMEIDA, 2011, such as reducing the number of experiments or unnecessary repetitions to obtain information from a system, and improvements in the quality of information obtained through the results; thus, promoting a significant reduction in labour and, consequently, reduction in the time and the final cost of completion experiments; optimization of more than one response at the same time, also allowing to calculate and evaluate the experimental error, and the possibility to be applied by competent researchers in its area without the need for extensive prior knowledge in statistics.

The full factorial design is generally used in early stages of a study (called screening) aiming to select the experimental variables and their interactions that mostly influence the response of interest (SOUSA et al., 2015).

In this type of planning, each element or factor, acting alone or interacting with other factors involved in the system, can be quantified and evaluated; being possible to infer, therefore, on the viability and quality of the final product developed (BUTON, 2012).

In many cases, it is possible to establish reliable conclusions from a well-designed experiment, using just enough elemental analysis techniques; however, in poorly designed experiments, even the most sophisticated statistical analysis cannot provide trustworthy conclusions (COSTA; ALMEIDA, 2011).

However, it should be noted that the use of softwares without prior knowledge of the methodology fundamentals, may represent a great risk, which may lead the user to dangerous misinterpretation of the data. By using these systems, the results are easily obtained, which is an excellent feature for conscious professionals, but for those who skip the fundamentals may present a great source of misinterpretations (COSTA; ALMEIDA, 2011).

PLANING THE EXPERIMENTS

When planning the experiments, the first thing to decide, is which factors should be studied, this means, the variables that can be controlled in the experiment, and which are the responses of interest. It is also necessary to specify the levels at which each factor should be studied, in other words, their values or classes that will be used in the experiments, because it is from that variations of levels that may or may not be produced variations in the response (BARROS NETO ; SCARMINIO; BRUNS, 2010).

A factorial design in which all variables are studied in just two levels is the most simple and, therefore, having k factors, the design of two levels will require the performance of 2 x 2 x 2 ... x = 2k different assays, and so being called factorial design 2^k (NORIEGA et al., 2005; BUTON, 2012).

Having selected the most important factors to be studied, the next step should be to quantitatively evaluate its influence on the response of interest, as well as the possible interactions amongst the factors (COSTA; ALMEIDA 2011).

The factorial design determines which factors have significant effects on the response, and also, how the effect of one factor varies with the levels of other factors. It also allows to establish and quantify correlations between different factors (CUNICO, 2008).

It is noteworthy that in factorial design is important to conduct repeated tests to estimate the experimental error, because the extent of this error will determine whether the effects observed were indeed significant. In the planning of two levels, it is often to arbitrarily identify the upper and lower levels with the signs (+) and (-), respectively, which points the necessity to carry out tests on all possible levels combinations (BARROS NETO; SCARMINIO; BRUNS, 2010).

To perform the factorial design, it is necessary to perform experiments and record the responses obtained in all possible combinations, being the list of these combinations called planning matrix, as shown in Tables 1 and 2 below.

Table 1. Factorial planning matrix 2^2

Assay	**Factor (1)**	**Factor (2)**	**Respons e**
1	**-**	**-**	______
2	**+**	**-**	______
3	**-**	**+**	______
4	**+**	**+**	______

Table 2. Factorial planning matrix 2^3

Assay	**Factor (1)**	**Factor (2)**	**Factor (3)**	**Response**
1	**-**	**-**	**-**	______
2	**+**	**-**	**-**	______
3	**-**	**+**	**-**	______
4	**+**	**+**	**-**	______
5	**-**	**-**	**+**	______
6	**+**	**-**	**+**	______
7	**-**	**+**	**+**	______
8	**+**	**+**	**+**	______

After selecting the factors and their levels to be studied, the tests should be performed in all possible combinations, according to the plans demonstrated in the factorial design matrices, remembering that the order of each assay should be done randomly, not being related to the number assigned to each assay, for example, assay 1, corresponding to factors at levels (- - -), should not necessarily be the first one to be performed.

To be used in the proposed tool, the tests must be performed in duplicate in order to estimate experimental errors, and thenceforth, determine the statistical significance of the effects. According to BARROS NETO; SCARMÍNIO; BRUNS, 2010, in the absence of any authentic repetitions in each experiment stage, the errors may seem smaller than they actually are, or perhaps significance is given to unreal effects.

For CUNICO, 2008, in a factorial design, the replicates or repetitions of experiments are fundamental to determine experimental errors in the response and/or the reproducibility of the experimental scheme used, thus, all assays and replicates must be performed randomly and authentically, aiming to avoid statistical distortions that compromise the quality of the results obtained and the effects calculated for the studied variables.

After performing the assays and obtaining the response observed in each experiment, the analysis through factorial planning determines which factors have relevant effects on the response, and also, how the effect of one factor varies with the levels of others. In addition, it allows to establish and quantify the correlations between the different factors (CUNICO et al., 2008).

In order to perform such calculations, the following tools have been developed using MICROSOFT OFFICE EXCEL®2007 SOFTWARE.

USING THE DATA ANALYSIS TOOL

The analysis of the data obtained in the factorial planning can be done using the tool described below, presented through Tables 3 and 4.

Table 3. Spreadsheet model for calculating the factors effects on the response under study and evaluation of the statistical significance of these in factorial planning 2^2.

	Factor 1	Factor 2	Replica 1	Replica 2	**Influences**	Values
1	-	-			**FACTOR 1**	0
2	+	-			**FACTOR 2**	0
3	-	+			**INTERACTION**	0
4	+	+			**SIG. LEVEL**	0

Source: Personal file

The original worksheet can be accessed through the link: https://docs.google.com/spreadsheets/d/1PCzYc5BiAXyZaAbmMNt6rBAJIUijk53YekLd0gu3wyo/edit#gid=1594036002 where the cells can be filled according to the answers obtained by the researcher for the response analysis in a factorial design 2^2.

The spreadsheet should be filled according to the assays performed, where each result obtained in each test must be expressed in its corresponding cell, respecting the definition of each experiment through the factorial planning spreadsheet 2^2. Afterwards, the values column, that corresponds to the influences of each factor on the result in question as well as the limit of significance for each effect will show the effects observed and their limit of significance from the calculations performed automatically.

The result shown for each effect represents the influence that each factor promoted in isolation, on average, in the studied response. The interaction amongst the studied variables can also be observed in the column.

Thus, the data can be discussed by observation of effects that had influence on the response in question, if there was interaction between them and if they were statistically significant or just experimental errors. Only effects with absolute values above the significance level, also expressed in the column, are considered statistically significant with a 95% confidence level, $p < 0.05$.

Table 4. Spreadsheet model for calculating the factors effects on the response under study and evaluation of the statistical significance of these in factorial planning 2^3.

Assay	Factor 1	Factor 2	Factor 3	Replica 1	Replica 2	Influences	Values
1	-	-	-			FACTOR 1	0
2	+	-	-			FACTOR 2	0
3	-	+	-			FACTOR 3	0
4	+	+	-			INT 12	0
5	-	-	+			INT 13	0
6	+	-	+			INT 23	0
7	-	+	+			INT 123	0
8	+	+	+			SIG. LEVEL	0

Source: Personal file

For the response analysis in a factorial design 2^3, the original spreadsheet can be accessed through the link https://docs.google.com/spreadsheets/d/1-_CALFZAvihL6OUxSimfr9HNIesEx-JCOburQtesMeA/edit#gid=200158841 where the cells can be filled according to the responses obtained by the researcher.

To calculate the effects using a factorial design 2^3, the filling of the spreadsheet should occur as described in planning 2^2, however in this case, there are 8 tests performed in duplicate, where each result obtained must be expressed in its corresponding cell, respecting the definition of each experiment through the factorial planning spreadsheet 2^3. Then, the values column, corresponding to the influences of each factor on the result in question as well as the significance limit for each effect will show the observed effects and their significance level. However, in this case, more effects will be observed in relation to the previous planning.

The values' column will contain the influence that each factor promoted in isolation, on average, in the response studied and in this case, a greater number of double interactions between the variables studied could be observed in the column (interactions between factors 1 and 2, 2 and 3 or 1 and 3), the column also shows the value of triple interactions between the factors under study, where only effects with absolute values higher than the established for the limit of significance should be considered statistically significant with 95% confidence, $P < 0.05$.

The calculations of the factors' effect and the interactions between them with their respective standard errors are performed as cited by BARROS NETO; SCARMÍNIO; BRUNS, 2010, through the spreadsheets.

UNDERSTANDING THE CALCULATIONS

The main effects are the mean differences of effects at the two levels of each factor, that represents the difference between the mean response at the upper level and the mean response at the lower level of this factor. When the response associated to the change in level in one factor varies when another factor is manipulated, it is said that interaction occurred between these factors. The interaction effects can be calculated by the difference in the responses obtained by level variation of the factor in each level of the factor with which it interacts. In fact, since each represents the means differences, where half of the observations contribute to one of the means and the other half to the other mean, half of that difference is the representation of the interaction effect.

In factorial designs 2^2, from the authentic repetitions, in a given combination of levels, it is possible to estimate the experimental error in this combination, allowing it to take the variance of this pair of values as an estimation of the typical variance of the experimental procedure. If it is assumed that the variance is the same throughout the entire experimental region, it is possible to combine information from all the tests and obtain a more robust estimation, a F-test can be used to confirm this validity. To obtain this joint estimation, it is possible to use the average of all estimations, weighted by their respective degrees of freedom. As each effect is a linear combination of observations at two distinct points and it is assumed, because of the random order and the authentic repetition that they are statistically independent values and have the same population variance, dividing the result by two is obtained the variance associated with an effect and from the square root of this value, it is obtained the standard error associated with an effect.

From the standard error, it is possible to construct a confidence interval for the values of the effects from t-Student distribution, being that in practice, only those effects whose estimations exceed in absolute value, the standard error by the point in Student Distribution, because only then will the confidence exclude the 0 value.

As in the factorial planning 2^2, in planning 2^3, any effect can be interpreted as the difference of means, and from the authentic repetitions, in a given combination of levels, it is possible to estimate the experimental error in that combination and then calculate the joint estimation of variance, based on the average of all the estimations, weighted by their respective degrees of freedom. However, in this case, each effect is a linear combination of 8 values, calculated from 4 means, assuming that these values are independent and dividing the joint variance by four, it is obtained the variance associated to an effect. The root of this variance represents the standard error of an effect that can be used along with Student's t-distribution to calculate the significance level of effects, as in the previous case.

PRACTICAL EXAMPLES

Example 1:

In a hypothetical case, a pharmacy student decided to study the most efficient way to extract the constituents of a plant commonly used to prepare infusions.

As in infusions preparations the soft parts of the plant, in the case, leaves, are left in contact with water previously boiled and covered to allow the constituents extraction, he decided to verify if the agitation of the mixture during the process and the state of the leaves would influence the content of extracted substances.

Thus, he decided to carry out a factorial design 2^2 to study this process, defining that he would study the influence of agitation, evaluating extractions made with and without agitation and the state of the leaves, where he would use the leaves undivided or crushed. In order to evaluate the efficiency of the extractive process, he decided to evaluate the dry residue obtained from the extracts after the extraction.

He then assembled the following factorial planning matrix, as shown in Table 5, performing the experiments in duplicate and at random, as presented in each assay, obtaining the results described for the dry residue.

Table 5. Factorial design matrix 2^2 to study the influence of agitation and leaf condition on the percentage of dry residue obtained in an extraction process.

Assay	**Factor Agitation**	**Factor Leaf state**	Response (%) Dry residue	
1	**Without**	**Undivided**	0,4	0,6
2	**With**	**Undivided**	1,4	1,5
3	**Without**	**Crushed**	0,8	1,1
4	**With**	**Crushed**	1,9	1,7

Transferring this data to the factorial calculation tool 2^2 it was obtained the following results.

Table 6. Effect of agitation and state of leaves and their interaction on the percentage of dry residue obtained in an extraction process.

Assay	Factor 1	Factor 2	Replica 1	Replica 2	Influences	Values
1	-	-	0,4	0,6	FACTOR 1	0,9
2	+	-	1,4	1,5	FACTOR 2	0,4
3	-	+	0,8	1,1	INTERATION	-0,05
4	+	+	1,9	1,7	SIG. LEVEL	0,294439264

Considering these results he observed that the factor agitation was the one that mostly influenced the dry residue yield, where the extractions made under mechanical agitation led to a 0.9% higher dry residue obtainment, on average, in relation to the assays where there was no agitation .

The size of the parts used also resulted in changes in the extraction process, where the use of the crushed leaves led to an increase of 0.4% on average in relation to the tests with undivided leaves, he observed that there was no significant interaction between these two factors, once the value in module assigned to the interaction did not exceed the significance limit. The factors could be interpreted in isolation, seeing that when the level of one of the factors was modified, there was a change in the response, and there was no evidence that this increase depended on the levels of other factors in the experimental area investigated.

Thus he concluded that the extraction made under agitation using crushed leaves resulted in a higher yield of dry residue.

Example 2:

Imagine now that this same student studied the previous process using a time of 20 min to carry out the extraction, however he also decided to study, in addition to how the extraction was made (with or without agitation) and the state of leaves (undivided or crushed), whether the extraction time would also influence the dry residue obtained.

So he decided to carry out a 2^3 factorial design to study this process, defining that he would study beyond the other two factors, if the extraction done in 20 minutes would have advantages over extractions made in only 10 min.

Table 7. Factorial planning matrix 2^3 to study the influence of agitation, leaf condition and extraction time on the percentage of dry residue obtained in an extraction process.

Ensaio	Factor agitation	Factor Leaf size	**Factor Time (min)**	**Response**	
1	Without	Undivided	**20**	0,4	0,7
2	With	Undivided	**20**	1,3	1,5
3	Without	Crushed	**20**	0,8	1,2
4	With	Crushed	**20**	1,9	1,8
5	Without	Undivided	**10**	0,4	0,5
6	With	Undivided	**10**	1,2	1,3
7	Without	Crushed	**10**	0,6	0,7
8	With	Crushed	**10**	1,7	1,8

The factorial design matrix was produced as shown in Table 6, performing the experiments in duplicate and at random, as presented in each assay, obtaining the results described for the dry residue.

Transferring this data to the factorial planning tool 2^3 it was obtained the following results:

Table 8. Effect of agitation, leaf condition, extraction time and their interactions on the percentage of dry residue obtained in an extraction process.

Assay	Factor 1	Factor 2	Factor 3	Replica 1	Replica 2	Influences	Values
1	-	-	-	0,4	0,7	FACTOR 1	0,9
2	+	-	-	1,3	1,5	FACTOR 2	0,4
3	-	+	-	0,8	1,2	FACTOR 3	-0,175
4	+	+	-	1,9	1,8	Int 12	0,075
5	-	-	+	0,4	0,5	Int 13	0,05
6	+	-	+	1,2	1,3	Int 23	-0,05
7	-	+	+	0,6	0,7	Int 123	0,075
8	+	+	+	1,7	1,8	Sig Level	0,168077188

Based on these results, he observed that agitation was the factor that mostly influenced the yield of dry residue, where the extractions made under mechanical agitation to a 0.9% higher dry residue obtainment, on average, in relation to the tests where there was no agitation.

The size of the leaves crushed also led to alterations in the extractive process, where the use of crushed leaves resulted in an increase of 0.4% on average in relation to the tests with undivided leaves.

Regarding the time of extraction, he observed that when using 10 min to perform the extraction, there was a decrease of 0.175% on average in the content of dry residue. However, the student concluded that this gain of 0.175% in the yield of dry residue could not be completely desired, upon extraction for a period of 20 min, since the infusion would take twice as long to be performed. The student also observed that there was no significant interaction between the factors, since the value in module attributed to the interaction did not exceed the limit of significance, which allows the factors to be interpreted in isolation.

Thus, he concluded that the extraction made under agitation using crushed leaves during 20 min obtains the best yield of dry residue. However, the extraction for a period of only 10 min under the same conditions also obtains satisfactory values, which could make the additional time spent on extraction less advantageous.

CONCLUSION

From data obtained through the assays of factorial design 2^2 and 2^3, conducted in duplicate and at random, using the developed tool, it is possible to quickly obtain the values of the effects of each factor on the response in study, and their interactions, also assessing the statistical significance of each one of them on the response in question, which allows a quick and efficient evaluation of the answers provided, facilitating the obtainment of these results. Thus, through a good experiment planning and use of the presented tools, it is possible to understand better the process studied, obtaining quick and practical results in the data analysis.

REFERENCES

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **Como fazer experimentos.** 4. ed. Porto Alegre, Bookman, 2010. 414 p.

BUTTON, S. T. **Metodologia para planejamento fatorial e análise de resultados**. Apostila: Programa de Pós-graduação em engenharia mecânica. Universidade Estadual de Campinas. Campinas, 2012.

COSTA, L. M.; KORN, G. A.; NOGUEIRA, A. R. A. Planejamento fatorial aplicado à digestão de amostras de feijão assistida por radiação microondas. **Quim. Nova,** v. 29, n. 1, p. 149-152, 2006.

COSTA, M. C. B.; ALMEIDA, C. R. R.; A importância do ensino da ferramenta de planejamento fatorial para estudantes de engenharia. XXXIX Congresso Brasileiro de Educação em engenharia -COBENGE, Blumenau-SC, 03 a 06 out de 2011.

CUNICO, M. W. M.; CUNICO, M. M.; MIGUEL, O. G.; ZAWADZKI, S. F.; PERALTA-ZAMORA, P.; VOLPATO, N. Planejamento fatorial: Uma ferramenta estatística valiosa para a definição de parâmetros experimentais empregados na pesquisa científica. **Visão Acadêmica**, Curitiba, v.9, n.1, 2008.

NORIEGA, P., RÖPKE, C. D., CAMILO, C. M., DEODATO DE FREITAS, P. C., BARROS, S. B. M. Avaliação por análise fatorial das condições da extração do 4- nerolidilcatecol de Pothomorphe umbellata (L). Miq. **Revista Brasileira de Ciências Farmacêuticas.** v. 41, n. 2, 2005.

PEREZ, T. A.; FERREIRA, N. N.; PREZOTTI, F. G.; GREMIÃO, M. P. D. Planejamento fatorial como ferramenta no desenvolvimento de complexos polieletrolíticos (PECs) a serem utilizados como sistemas de liberação modificada. **Rev. Ciênc. Farm. Básica Apl.** v.36, 2015.

SOUZA, M. O.; RAINHA, K. P.; CASTRO, E. V. R.; CARNEIRO, M. T. W. D.; FERREIRA, R. Q. Análise exploratória das concentrações dos metais Na, Ca, Mg, Sr e Fe em extrato aquoso de petróleo, determinados por ICP OES, após otimização empregando planejamento de experimentos. **Quím. Nova**. v.38, n.7, 2015.