# FUNDAMENTAL CONCEPTS AND RECENT APPLICATIONS OF FACTORIAL STATISTICAL DESIGNS

Pedro Henrique MAINARDI[1]
Ederio Dino BIDOIA[1]

- ABSTRACT: Factorial designs have been increasingly used in scientific investigations and technological development. The designs, through the use of matrices with all the treatment combinations, have been capable to effectively characterize the relationships between the variables of multi-factor experiments, assess the experimental variabilities, and derive mathematical functions that represent the behavior of the responses. Factorial designs were fractionalized, which substantially reduced the number of treatments without the loss of relevant information. The addition of central and star points to the factorial arrays has given them the orthogonality and rotatability characteristics, frequently used to fit models with curvature and identify critical regions of interest. Literature reports indicated that factorial designs, also called factorial experiments, were successfully applied in different types of investigations, including in cost evaluations and time-series studies. They were capable to estimate important features of the experiments, like the individual and combined effects of factors, the magnitude of residuals, additionally to express the relationships of the variables in polynomial equations, draw response surface and contour plots, and determine optimal combinations of parameters. In this review, the fundamental aspects of the Complete, Fractional, Central Composite Rotational and Asymmetrical factorial designs were conceptualized, and recent applications of these powerful tools were described.

- KEYWORDS: Design of experiment; DOE; experimental planning; process optimization; research design; statistical plan; statistical model.

## 1 Introduction

Experimentation is an essential part of the scientific methodology, fundamental for the production of knowledge, progress of science and technological development (ZANCAN, 2000). The process of experimentation, initially defined as a "deliberate observation under conditions deliberately arranged by the observer", has been summarized to a "a cause-and-effect relationships in a system", an essential requirement to demonstrate

[1]Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, Instituto de Biociências, Departamento de Biologia Geral e Aplicada, CEP: 13506-900. Rio Claro, SP, Brazil, E-mail: pedro.h.mainardi@unesp.br; ederio.bidoia@unesp.br

theories or hypotheses of a functional structure (HINKELMANN and KEMPTHORNE, 2007; MONTGOMERY, 2017). The application of statistical designs in experimentation, which now has been a common practice in academic laboratories to ensure valid conclusions in objective investigations, has begun in the 1930s, when the activities of getting products with better quality and processes with higher productivity had become a goal in industries (MASON *et al.*, 2003; MONTGOMERY, 2017). Since then, the experimentation process through statistical designs has been increasingly used, along to achieve products with better quality within the limits of desired specifications, to substantially reduce the development time and cost, promote reliable and environmental-friendly processes, and to proceed experimental analyses and data collection more efficiently (MASON *et al.*, 2003; RODRIGUES and IEMMA, 2005; WEISSMAN and ANDERSON, 2015; MONTGOMERY, 2017).

The use of statistical designs, also called statistical plans, has enabled the researchers to, within the rigor and adherence to the principles of science, generate the maximum amount of relevant information in a minimum number of possible experiments (MASON *et al.*, 2003). They have been used to accurately draw inferences from the observational data, characterize the relationships between different types of experimental variables, and economically assess the relevance and the quantifying effects of factors (CHRISTENSEN, 1996; MASON *et al.*, 2003). Different from OFAT (one-factor-at-a-time) experiments that only shows how a response variable is affected by varying each factor while all the other are kept constant, well planned statistical designs, like the factorial ones, have enabled the consideration of any possible interaction between the factors, a remarkable achievement in modern experimentation (MONTGOMERY, 2017). Factorial designs, in which the factors are varied together instead of one at a time, have been an important approach for scientists and designers to, without the excessive use of experimental resources and waste through uneconomical methods, quantify synergistic or antagonistic effects of factors and estimate ideal combinations of variables (MASON *et al.*, 2003; RODRIGUES and IEMMA, 2005).

Another great feature of statistical designs is that they have been able to measure the magnitude of errors, that is, random variations that can occur due to changes in the ambient condition, experimental and measurement errors, or effects due to any other known or unknown influences (MASON *et al.*, 2003). The possibility to distinguish assignable causes of variations from random ones has been widely used to eliminate sources of bias (systematic differences) and ensure that the experiment has provided precise information concerning the responses of interest (MASON *et al.*, 2003). They have been successfully able, thus, to properly measure the relationships between the variables and the influence of experimental factors (DRAPER and SMITH, 1998; HINKELMANN and KEMPTHORNE, 2007). Additionally, they have been successfully capable to derive simple and complex polynomial equations regarding the behavior of the experimental data, draw response surface and contour plots that simplify the interpretation of the empirical results, forecast future events based on previously observed values, and determine optimal operational conditions of different kind of systems (MASON *et al.*, 2003; RODRIGUES and IEMMA, 2005).

Due to the growing importance that statistical designs have been acquiring in modern researches, as well the great relevance that they have been detaining in scientific and technological areas, this review article addressed the conceptualization of their fundamentals and described the characteristics of the Complete, Fractional, Central

Composite Rotational and Asymmetrical factorial designs. The article also addressed the sequential use of factorial statistical designs as an experimental optimization strategy, their employment in cost evaluations and time-series studies, and summarized recent applications of their use in several scientific fields.

## 2 Methodology

The literature review was based on physical documents obtained in the library of the local University and in virtual databases indexed on the Web of Science, Scopus, ScienceDirect and Google Scholar platforms. The surveys were conducted until August 12th, 2021, and included both public open-access and institutional available documents published in English or Portuguese. The terms used in the searchers were "Fundaments of Statistics", "Concepts of Statistics", "Statistical Analysis", "Design of Experiment", "Statistical Inference", "Confidence Interval", "Prediction Interval", "Regression Analysis", "Factor Effects", "Determination Coefficient", "Analysis of Variance", "ANOVA", "Lack-of-Fit", "Outlier", "Pareto Diagram", "Response Surface", "Contour Plot", "Multi-factor Design", "Factorial Design", "Complete Factorial", "Full Factorial", "Fractional Factorial", "Central Composite Rotational", "Asymmetrical Factorial", "Mixed-Level Factorial", "Sequential Use of Factorial", "Plackett-Burman Design", "Taguchi Design", "Cost Evaluation", "Economic Analysis", "Time-Series" and "Time-Trend". After the acquirement of the documents, they were screened according to their titles, abstracts and contents in order to eliminate duplicates and verify their adequacy to the theme proposed in this article. Then, the documents were classified according to the following topics: (i) Fundaments of Statistics, (ii) Design of Experiments, (iii) Complete Factorial, (iv) Fractional Factorial, (v) Central Composite, (vi) Asymmetrical Factorial, (vii) Sequential use of Factorial Designs, (viii) Cost Analyses/Evaluations and (ix) Time-Series Studies.

## 3 Results

The literature survey found 216 documents that had adherence to the theme proposed by this review article. After the screening process mentioned in the previous item, 66 documents were classified as belonging to the topic (i), 28 as belonging to the topic (ii), 12 to the topic (iii), 15 to the topic (iv), 20 to the topic (v), 13 to the topic (vi), 16 to the topic (vii), 17 to topic (viii) and 29 to the topic (ix). Of these, 98 documents that presented great quality were chosen to be used as a theoretical basis in the preparation of this article and included in the references.

## 4 Principles of statistical designs

Analyze and understand a relevant set of data from a certain investigation are frequent practices in the research activity of different study fields. Descriptive statistical values, which summarize datasets, such as the simple arithmetic mean, mode, median, and dispersion measures, like the amplitude, deviations and variances, have been notoriously

used to explore and simplify the interpretation of empirical observations (Equations 1, 2, 3, 4, 5 and 6) (PEARSON, 1894; YULE, 1897; FISHER, 1919; ZWILLINGER and KOKOSKA, 1999; ALTMAN and BLAND, 2005; SHESKIN, 2011). Regularities or patterns identified from the observations, though, have enabled the researchers to transform the data into more useful information, a fundamental approach to draw valid inferences from the observational data and develop models that adequately describe them (MASON *et al.*, 2003; HINKELMANN and KEMPTHORNE, 2007). Statistical analyses, in this context, have been playing a major role to judge the adequacy of scientific hypotheses, compare results in the literature, and make strategic decisions (BROWNLEE, 1965; MORETTIN and BUSSAB, 2017).

The possibility to develop experimentation designs, that is, statistically designed experiments that allows an efficient measurement of the relationship among variables of interest and the estimation of the magnitude of experimental errors, have enabled the researchers to comprehend cause-and-effect relations in a system, thus, represent a theory and make accurate forecasts based on their results (MASON *et al.*, 2003; KUTNER *et al.*, 2005; MONTGOMERY, 2017). These designs, where the variables of interest are often controlled and fixed at predetermined values for each test run, have been named statistical designs, also known as statistical experiments (MASON *et al.*, 2003). Statistical designs, besides developed from empirical observations obtained in experimental studies, have also considered various types of statistical parameters, specifications and measurement errors (MASON *et al.*, 2003). The designs' assumptions, which can have many forms and configurations, have been mainly confirmed from adequation proceedings, adjustment measures or subjectively by graphic techniques (CHRISTENSEN, 1996; MASON *et al.*, 2003).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{3}$$

$$SE = \frac{\sigma}{\sqrt{N}} \quad \text{or} \quad \frac{S}{\sqrt{n}} \tag{4}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{5}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{6}$$

where $\bar{x}$ refers to the arithmetic mean, $x_i$ to the response of a random variable, $n$ to the sample size, $S$ to the sample standard deviation, $\sigma$ to the population standard deviation, $\mu$ to the population mean, $N$ to the population size, $SE$ to the sample or populational standard error, $S^2$ to the sample variance, and $\sigma^2$ to the population variance.

## 5  Statistical inferences

Statistical designs have enabled the estimation of conclusions about certain population characteristics and processes through measurements or observations made in a representative set, called the sampling group (CHRISTENSEN, 1996; MORETTIN and BUSSAB, 2017). The conclusions, also known as statements or inferences, have involved a degree of uncertainty due to uncontrollable experimental variations based on statistical models that represent the probability of the occurrence of events, named probabilistic frequency distributions (GALTON, 1889; MASON *et al.*, 2003; ONYIAH, 2008). Most of statistical designs have been assumed to detain a Normal or close to Normal frequencies, that is, when the data is symmetrically concentrated around a central value. Statistical designs with small sample sizes (n < 30), however, in order to adjust the number of observations through the degree of freedom, have commonly replaced the Normal distribution by the *Student*'s t distribution, and the populational standard deviations by the sample deviations (RODRIGUES and IEMMA, 2005; ONYIAH, 2008).

The assumptions of the inferences have been done by the use of hypothesis tests, which uses a validation criterion with two suppositions: a null ($H_0$) and an alternative hypothesis ($H_A$) (CHRISTENSEN, 1996). The null hypothesis has been commonly used to state that there was sufficient evidence to prove that there was no relationship between two measured parameters, and the alternative hypothesis, that there was not enough evidence for such statement. The results of the tests have been compared to tabulated critical regions of the Normal or the *Student*'s t distributions models according to determined levels of significance (Equations 7 and 8) (MORETTIN and BUSSAB, 2017). The levels of significance, also called α value, have been used to determine the probability to mistakenly reject the null hypothesis when it was true, that is, the probability of the occurrence of a false positive (KALINOWSKI and FIDLER, 2010). The test uses a measurement of evidence related to the probability to obtain extreme results under the null hypothesis, called *p*-value, probability value or descriptive level (BROWNLEE, 1965). The *p*-value has been interpreted as a decreasing index of reliability, that is, the higher its value, the less reliable has been the observed relationship between the parameters of interest.

$$Z_{calc} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{7}$$

$$t_{calc} = \frac{\bar{x} - \mu}{S / \sqrt{n}} \tag{8}$$

where $Z_{calc}$ and $t_{calc}$ refers to the test statistics following the respective Normal or *Student*'s t distributions, $\bar{x}$ to the sample mean, $\mu$ to the population mean, $n$ to the number of

independent observations (*Student*'s *t* distribution with a *n – 1* degree of freedom), $\sigma$ to the population standard deviation, and *S* to the sample standard deviation.

## 6  Confidence and prediction intervals

Confidence intervals, a type of estimate that infers information regarding the precision of the estimators, have been used to represent repetition frequencies around estimates of uncertainties and to indicate the inaccuracy of the means (NAKAGAWA and CUTHILL, 2007; PATINO and FERREIRA, 2015). In statistical designs, the intervals are associated to a confidence level whose amplitude estimates the frequency in which the value of a parameter of interest would be included in infinite hypothetical repetitions (KALINOWSKI and FIDLER, 2010). Confidence intervals of 95%, for example, have indicated that the mean value of the parameter of interest would tend to be included within the calculated extremities 95% of the time if the procedure was repeated numerous times (COX and HINKLEY, 1979). The prediction interval has also been interpreted as an interval estimate calculated from observations, however, it has indicated values in which a new observation would probably fall according to a certain probability. This type of interval has provided more accurate estimates in relation to the occurrences of events and experimental responses (CHRISTENSEN, 1996). The calculation of a 100(1-α)% confidence interval of a normally distributed large sample size with a known populational standard deviation has been done as described in equation 9. The prediction interval has been calculated as indicated in equation 10 (ONYIAH, 2008).

$$\hat{\theta} = \bar{x} \mp z_\alpha \frac{\sigma}{\sqrt{n}} \tag{9}$$

$$\hat{\theta} = \bar{x} \pm z_\alpha \sigma \tag{10}$$

where $\hat{\theta}$ refers to the interval estimate, $\bar{x}$ to the sample mean, *Z* to the tabulated value of the Normal distribution, α to the probability of the confidence level, $\sigma$ to the population standard deviation, and *n* to the number of independent observations.

## 7  Regression analyzes

Statistical designs, through regression methods, have enabled the characterization of the relationships between independent and dependent variables used in the experiments, so called predictor and response variables respectively (DRAPER and SMITH, 1998; MONTGOMERY, 2017). Regression techniques, widely used in statistical modeling, besides used to quantitatively estimate the contribution of independent variables in the prediction of the dependent ones, have been used to assess the adequacy of the assumptions, examine the influence of atypical observations (outliers), and also to identify the presence of linear effects between variables, called correlations (CHRISTENSEN, 1996; MASON *et al.*, 2003). In general, the relationships between the variables have been expressed in

mathematical functions calculated using the least squares method (DRAPER and SMITH, 1998; ZWILLINGER and KOKOSKA, 1999).

The least squares method has provided a numerical estimation of the vertical intercepts and the slopes of the regression equations, in other words, the values of the constants and the angular coefficients, also called gradient of the linear functions. The associations of the responses have been described through simple or complex linear coefficients, that is, of first or more orders (Equations 11 and 12). The regression models have been referred as linear because the unknown coefficients have appeared in linear forms, that is, as additives or multipliers constants of the values of the predictor variables (MASON *et al.*, 2003). Regression models have been diagnosed through adjustment measures, inferences analysis, and visually through graphics, such as dispersions and scatterplots (CHRISTENSEN, 1996; MASON *et al.*, 2003; MONTGOMERY, 2017).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e \tag{11}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e \tag{12}$$

where *y* refers to the predicted response of the adjusted model, $x_{...}$ to the predictor variables, $\beta_0$ to the intercept value, $\beta_{...}$ to the simple or quadratic (parabolic) linear coefficients, *k* to the number of predictor variables, and *e* to the adjustment error.

## 8 Factors' effects

In statistical designs, the independent variables, also referred as the predictors or controllable variables, have been commonly called factors. As seen in regression models, the relationships between the factors and the responses have been quantitatively calculated and described in the form of polynomial functions (Equation 13) (BEZERRA *et al.*, 2008). The effects of the factors have also been used in inferences tests and interval estimates (MASON *et al.*, 2003). The interactions and correlated effects, non-linear and linear relationships between the responses of three or more variables, have occurred when the results of one factor depended on the state of another factor. The interactions effects have also been measured and described in the form of mathematical equations in regression models (Equation 14) (CHRISTENSEN, 1996). In balanced statistical designs, that is, with equal amounts of repetitions for all the combinations of equally spaced factor's values, the coefficients of the effects have been previously tabulated, which had simplified the derivation of polynomial functions of second or more orders (Equation 15) (MASON *et al.*, 2003). The values of the predictor variables have been used in a standardized and coded forms, facilitating comparative analyses (Equation 16).

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + e \tag{13}$$

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{1 \le i \le j}^{k} \beta_{ij} x_i x_j + e \tag{14}$$

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{1 \leq i \leq j}^{k} \beta_{ij} x_i x_j + e \qquad (15)$$

$$z_i = \frac{x_i - \bar{x}_i}{S_i} \qquad (16)$$

where $y$ refers to the predicted response, $k$ to the number of variables, $\beta_0$ to the constant term, $\beta_i$ to the simple linear coefficient, $\beta_{ij}$ to the interaction coefficient, $\beta_{ii}$ to the quadratic coefficient (second order), $x_i$ and $x_j$ to the predictor variables, $e$ to the residue, $z_i$ to a standardized predictor variable, $\bar{x}_i$ to the sample mean, and $S_i$ to the sample standard deviation.

## 9  Coefficients of determination

The coefficient of determination, also called explanatory coefficient or $R^2$, has been interpreted as a measure of the percentage of responses that was explained by the regression equations. In other words, the determination coefficient has quantified the proportion of the adequacy between the observed values in relation to the proposed values by the statistical design (RODRIGUES and IEMMA, 2005; NAKAGAWA and SCHIELZETH, 2013). The adjusted coefficient of determination ($R^2a$), in turn, has been defined as an adjustment of the $R^2$, as it has also considered the number of independent observations, the degree of freedom. The adjusted determination coefficient has been used to compare statistical designs with different numbers of predictor variables (CHRISTENSEN, 1996). The quality of the regression equation fit, according to the determination coefficient $R^2$, has been calculated as indicated in equation 17. The calculation of the adjusted $R^2$ has been done as demonstrated in equation 18. In general, $R^2$ and $R^2a$ values have varied between 1 and 0, and have been expressed as percentages.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad (17)$$

$$R^2a = 1 - \frac{n-1}{n-k-1} (1 - R^2) \qquad (18)$$

where $R^2$ refers to the determination coefficient, $R^2a$ to the adjusted determination coefficient, $y_i$ to the experimental value observed in the response variable as a function of the level $i$ of a predictor variable, $\bar{y}$ to the general mean of the observed values, $\hat{y}_i$ to the estimated or predicted value of the adjusted model for a response variable as a function of the level $i$ of a predictor variable, $n$ to the number of observations, and $k$ to the number of predictor variables.

## 10 Analysis of variance

Through the use of analytical techniques around the variances, statistical designs have been capable to estimate significant differences between the means of treatment groups. The method, called ANOVA, has enabled to separate or partition the observable variation of the designs into two components: assignable and uncontrollable causes of variations, and suppose if there were significant differences between them (MASON *et al.*, 2003). The assignable causes have been considered as sources of variations that were possible to be measured and controllable, while the uncontrollable, sources of random variations (RODRIGUES and IEMMA, 2005; MORETTIN and BUSSAB, 2017). In general, the variability decomposition of random and independent samples, with normally distributed frequencies and equal populational variances, has been done through the Total Sum of Squares (Equations 19 and 20). The Sum of Squares of the Error and the Sum of Squares of the Regression model, respectively called residue and adjusted model, have been calculated using the equations 21 and 22.

$$TSS = SSR + SSE \tag{19}$$

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{20}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{21}$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{22}$$

where *TSS* refers to the sum of squares due to Total variation, *SSR* to the sum of squares due to parameters considered in the model (Regression), *SSE* to the sum of squares due to adjustment Errors (residue), $y_i$ to the value observed in the response variable as a function of the level *i* of a predictor variable, $\bar{y}$ to the general mean of the observed values, $\hat{y}_i$ to the predicted value of the adjusted model for a response variable as a function of the level *i* of a predictor variable, and *n* to the number of independent observations.

## 11 ANOVA significance tests

In analyzes of variance, the probability to obtain conclusive statements about the differences among the means of treatment groups, in general, have been calculated through the Mean Square of the Regression and the Residue ratio, called *F* test (Equation 23). The estimates of the test have been compared to critical regions of the *Fisher-Snedecor* distribution according to previously determined significance levels (HOAGLIN *et al.*, 1991; CHRISTENSEN, 1996). The ANOVA test has been commonly used to measure the magnitude of the statistical significance, that is, the estimates related to the "degree of certainty" between the associations of the parameters. ANOVA procedures have been frequently used in multiple variables analysis, where the *F* test has been done independently

for each factor and interaction, including covariates, blocks and curvatures (SHESKIN, 2011).

$$F_{calc} = \frac{MSreg}{MSres} \tag{23}$$

where $F_{calc}$ refers to the value of the test statistic following the $F$ distribution, *MSreg* to the mean square of the Regression, and *MSres* to the mean square of the Residue.

## 12 Lack-of-fit tests

The residuals of statistical designs are important parameters of reliability. They have been commonly used in inferences methods that allow the validation of the adequacy of the hypothetical models (RODRIGUES and IEMMA, 2005). The estimates have been done by the use of the variation between repetitions within each treatment, that is, random fluctuations between authentic replicates (DRAPER and SMITH, 1998; BEZERRA *et al.*, 2008). The calculations have been made by partitioning the Sum of the Squares of the Residues into two useful components for a hypothesis test according to certain levels of significance, so called Pure Error and Lack of Fit (Equations 24 and 25). Statistically significant lack-of-fit hypothesis tests ($p \leq \alpha$) have, in general, indicated a residual variability greater than a pure error variability (Equation 26).

Depending on the significance criterion, significant lack-of-fit tests have indicated a low degree of model adequacy, requiring, in some cases, the use of more complex designs to describe the behavior of the responses (RODRIGUES and IEMMA, 2005). Noteworthy, the lack-of-fit test by the sum of squares has been highly sensitive regarding deviations from normality, large or small sample sizes, and the presence of influential observations, those with atypical values, called outliers (HOOPER *et al.*, 2008). During the past years, remarkable researches have been done involved in the development of new diagnostic techniques and structural transformations of statistical designs. Methods, such as goodness-of-fit, incremental fit indices, parsimony fit indices, among others, have been proposed to validate the assumptions of different types of models (D'AGOSTINO, 1986; MULAIK *et al.*, 1989; STEIGER, 1990; MILES and SHEVLIN, 2007; HOOPER *et al.*, 2008; MONTGOMERY, 2017).

$$SSPE = \sum_{i=1}^{n}(y_{ij} - \bar{y}_i)^2 \tag{24}$$

$$SSRes = SSLOF + SSPE =$$

$$\sum_{i=1}^{n}(y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{n}(\bar{y}_{ij} - \hat{y}_i)^2 + \sum_{i=1}^{n}(y_{ij} - \bar{y}_i)^2 \tag{25}$$

$$F_{LOFcalc} = \frac{SSLOF}{SSPE} \tag{26}$$

where *SSPE* refers to the sum of squares of the Pure Error, *SSRes* to the sum of squares of the Residue, *SSLOF* to the sum of squares of lack of fit, $F_{LOFcalc}$ to the value of the test statistic following the $F$ distribution, $y_{ij}$ to the observed value of the $j$ repetition in the

response variable as a function of the level $i$ of a predictor variable, $\bar{y}_i$ to the general mean of the values observed in the level $i$ of a predictor variable, $\hat{y}_i$ to the value predicted through the adjusted model for a response variable as a function of the level $i$ of a predictor variable, $\bar{y}_{ij}$ to the mean value of observations between the repetitions at the level $i$ of a predictor variable, and $n$ to the number of independent observations.

## 13 Outliers

The presence of outliers, observations with atypical values in relation to other observations obtained under the same condition, was seen to strongly affect the inferences procedures and the regression models (MASON *et al.*, 2003). Outliers, usually with a high residual value, have been visually identified through graphs and diagrams, such as box plots, residual plots, normal *Quantil-Quantil*, and also by diagnostic techniques, such as Grubbs test, Leverage values, Cook's distance, standardized and studentized analysis of residues (GRUBBS *et al.*, 1950; COOK, 1977; CHRISTENSEN, 1996; KUTNER *et al.*, 2005; MONTGOMERY, 2017). Accommodation methods, such as the collection of more data, model reexpression, deletion, winsorizing and trimming extreme observations, have been often used to mitigate the effects of outliers (TUKEY, 1962; MASON *et al.*, 2003). The techniques, however, have been proceeded in a careful, justified and gradual way, based on stepwise selection methods, such as forward selection, backward elimination and stepwise iteration (CHRISTENSEN, 1996; DRAPER and SMITH, 1998; SHESKIN, 2011).

## 14 Pareto diagrams

Pareto diagrams, graphs that contain histograms and lines, have been used to order the frequencies according to their occurrences. In statistical designs, Pareto diagrams have been used to order the individual and combined effects of the variables according to the magnitudes of their statistical significance. The description of the effects in absolute and standardized values has enabled the inclusion of reference lines that have indicated a minimum magnitude for a significant statistical effect according to determined α levels of significance (RODRIGUES and IEMMA, 2005). In statistical modeling, Pareto diagrams have permitted the prioritization of variables of greater importance in a simplified way and frequently used in quality control charts (PORTER *et al.*, 1997).

## 15 Response surface and contour plots

The relationship between two or more factors has been drawn on graphs with two or three dimensions, called contour or response surface plots. The graphs, constructed through mathematical functions regarding the relationship between the independent and dependent variables, have been used to simplify the interpretation of empirical results, mainly in experiments with multiple factors (MASON *et al.*, 2003; HINKELMANN and KEMPTHORNE, 2007). In scientific investigations, contour and response surface graphs have been frequently used to evaluate important characteristics of statistical designs, such

as the behavior of the results, sensitivity of the variables and effects of interactions. In the assessment of industrial processes, response surface and contour charts have been used to improve the quality of products and to estimate optimal operational conditions of the systems (NOORDIN *et al.*, 2004; RODRIGUES and IEMMA, 2005).

## 16 Designs' equations

The behavior of experimental responses of statistical designs have been described in mathematical functions, expressed, in general, as demonstrated in equation 27 (MASON *et al.*, 2003). In the equation, the smooth or the regular part, so called *μ,* refers to the predictable part of the statistical design. The *e* part, considered as important as the *μ* part, refers to adjustment errors, also called residue (MORETTIN and BUSSAB, 2017). The residue is calculated from uncontrollable sources of variation, and has been interpreted as a measure of discrepancy between the observed and the proposed values by the statistical designs. Residual analyzes, which are usually done through adjustment measurements or by the use of graphs and diagrams, have been crucial to validate the adequacy of theoretical models (CHRISTENSEN, 1996; TSAI *et al.*, 1998).

$$y = \mu + e \qquad (27)$$

where *y* refers to the representation or the description of the observed data, *μ* to the statistical design, and *e* to the adjustment error (residual).

## 17 Multi-factor designs

Statistical designs have been increasingly used in experiments with multiple factors, also called multi-factors. The application of appropriate statistical designs in this type of study has allowed an efficient characterization of the individual and the interaction effects of the factors, in addition to provide the acquisition of precious estimates around the variability of adjustment errors (BEZERRA *et al.*, 2008; ANDERSON and WHITCOMB, 2010). Statistical designs with multiple factors have been commonly used in the construction of mathematical equations that describe the behavior of experimental responses in certain regions of interest (MASON *et al.*, 2003). The functions of first, second, third or more orders, derived from multi-factor statistical designs, have been used in the synthesis of response surface graphs and contour plots. The approach has enabled a quantifying estimation of the factors' effects, identification of critical regions, and optimization of responses (RODRIGUES and IEMMA, 2005; WEISSMAN and ANDERSON, 2015).

## 18 Complete factorial designs

The choice between the various types of statistical designs has depended on the objective, number of factors involved, available resources, and especially, the number of

required experiments (MASON *et al.*, 2003; WEISSMAN and ANDERSON, 2015). The Complete Factorial (CF) design, a classic multi-factor statistical design, has enabled all the treatment combinations through the use of a matrix with predetermined and equidistant coefficients (FISHER, 1936). CF designs, due to their broad applications in many research fields, have been increasingly used to characterize the individual and combined effects of different type of variables, estimate variations concerning the measurement errors, and also to derive graphs and mathematical functions that represent the behavior of experimental responses (HINKELMANN and KEMPTHORNE, 2007; SHESKIN, 2011; MONTGOMERY, 2017). Depending on the availability of time and experimental resources, the sequences of the factorial testing runs, that is, the assignments of each factor-level combination, have been capable to be minimally or maximally randomized by algorithms according to different criteria, which tend to decrease nonconstant variances, time trend effects and possible sources of bias (CHENG *et al.*, 1998; ANGELOPOULOS *et al.*, 2009; HILOW, 2013; BHOWMIK *et al.*, 2017). The treatment points of the matrix of a CF design, also called Full Factorial design, with three factors and three levels ($3^3$) were represented in Figure 1. The second order polynomial function generated by the regression model of a $3^3$ CF planning was described in equation 28.

According to reports, CF designs have been successfully used in the study of different types of processes, including the optimization of an eco-friendly biosorption process of a cotton dye, and the extraction method of a chemical compound with pharmacological activity from a plant's leaves (HENN *et al.*, 2019; MOGHAZY *et al.*, 2019). The authors stated that the CF designs were effectively used to determine the factors of the processes that were statistically significant. According to the descriptions, the polynomial equations built from the CF designs were successfully used to evaluate the individual and combined influences of the parameters, and provided robust assessments about the variables' effects as a function of the responses. The CF designs were also used by Porto *et al.* (2019) and Lara *et al.* (2019) to evaluate the performance of anticorrosive alloys, and to optimize the dielectric properties of a magneto-dielectric composite. According to the authors, the CF designs enabled an efficient investigation of the parameters that were used to achieve desired characteristics, and were successfully used to improve the performance of the systems. Ravindran *et al.* (2020) also used a CF design to screen out the significant constituents of a growth medium that was used for the cultivation of a freshwater microalga. The CF design was able to determine the multiple significant factors that were affecting the growth rate and the final biomass of the microorganism, as much as the existence of interactions between the variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 +$$

$$\beta_{122} x_1 x_2^2 + \beta_{112} x_1^2 x_2 + \beta_{1122} x_1^2 x_2^2 + \beta_{13} x_1 x_3 + \beta_{133} x_1 x_3^2 + \beta_{113} x_1^2 x_3 +$$

$$\beta_{1133} x_1^2 x_3^2 + \beta_{23} x_2 x_3 + \beta_{233} x_2 x_3^2 + \beta_{223} x_2^2 x_3 + \beta_{2233} x_2^2 x_3^2 + e \qquad (28)$$

where $y$ refers to the predicted response of the adjusted design, $\beta_0$ to the value of the intercept, $x_1$, $x_2$ and $x_3$ to the predictor variables, $\beta...$ to the partial linear coefficients, and $e$ to the adjustment error.
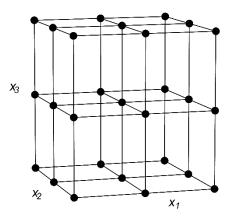


Figure 1 - Representation of a Complete Factorial (FC) statistical design with three levels and three factors ($3^3$).

## 19 Fractional factorial designs

Due to limitations involving time and availability of resources, Complete Factorial designs have not always been possible to be conducted. In such cases, Fractional Factorial (FF) designs have been used, that is, models that were consisted of a subset of the experimental points of a CF planning, a fraction of the CF treatment runs (BOX and HUNTER, 1961). The fractionalization of factorial designs has been made by the use of confounded effects, whose values could only be attributed to the combined influences of responses, and not to single individual responses, also called aliased effects (MASON *et al.*, 2003). The fractions of the FF matrices have been carefully chosen, ensuring that the effects of interest were not confounded with others of interest or, at least, were confounded with the effects of variables that did not have appreciable magnitudes (MASON *et al.*, 2003). This type of application has enabled a significant reduction of the number of experiments and, even so, has provided a comprehensive investigation of the factors without the loss of relevant information (BOX and HUNTER, 1961; GUNST and MASON, 2009). The treatment points of the matrix of a one-third FF design with three factors and three levels ($3^{3-1}$) were represented in figure 2. The second order polynomial function generated by the regression model of the FF planning was described in equation 29.

Reports indicated that FF designs have been successfully used to identify significant variables in screening scientific experiments and industrial processes. Studies by Pan *et al.* (2019), who used a FF design to develop and optimize a fermentation medium for the production of a biopolymer by a lactic bacterium, indicated that the experimental design

allowed the authors to simultaneously evaluate a large number of variables in a reduced number of experiments. Lim *et al.* (2020), in turn, also used a FF plan to optimize the combinations of the drugs that were used in the therapy of cancer cell lines. The authors reported that the FF statistical design enabled an accurately identification and optimization of effective factors combinations. FF designs were also used in the investigation of the impacts that various constituents had on the microbial activity of a plant growth media, and in the study of the effects that design parameters had on the performance of reinforced concrete bridge piers strengthened with steel-reinforced polymer composites (VAN GERREWEY *et al.*, 2020; WAKJIRA *et al.*, 2020). The FF design also allowed Rocha *et al.* (2021) to produce cost effective and efficient magnetic activated carbon for the adsorptive removal of pharmaceuticals from aqueous media, and Al-Dawalibi *et al.* (2020) to select the best marketing strategy for the purpose of increasing the sales revenue. The authors stated that the FF designs, in a reduced number of trials, enabled the identification of the most important experimental variables and were also extremely useful to investigate the main effects and possible interactions among them.

$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 +$$

$$\beta_{122} x_1 x_2^2 + e \tag{29}$$

where $y$ refers to the predicted response of the adjusted design, $\beta_0$ to the value of the intercept, $x_1$, $x_2$ and $x_3$ to the predictor variables, $\beta...$ to the partial linear coefficients, and $e$ to the adjustment error.
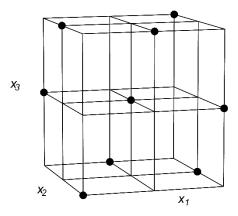


Figure 2 - Representation of a one-third Fractional Factorial (FF) statistical design with three levels and three factors ($3^{3-1}$).

## 20 Central Composite Rotational designs

The Central Composite Rotational (CCR) design, developed by Box & Hunter (1957), is consisted of a factorial design with orthogonality and rotationality (rotatability) characteristics, acquired by the insertion of star (alpha) and central points to the factorial arrangements. The conditions have provided independent estimates for the model coefficients and identical variances for all the treatment points, situated at the same distance from the center (CONAGIN, 1982). CCR designs have enabled the obtention of robust experimental information in a relatively modest number of treatments, including the quantification of individual and combined effects of the factors, the estimation of variations related to procedural errors, and also the construction of equations that express responses with curvatures, that is, with second-order functions (MASON *et al.*, 2003; ONYIAH, 2008). The adjusted models of CCR designs have been often used to economically synthesize contour plots and response surface graphs (RODRIGUES and IEMMA, 2005). The treatment points of the matrix of a CCR design with three factors, two levels and six-star points, were represented in figure 3. The second order polynomial function generated by the regression model of the CCR planning was described in equation 30.

According to reports, a Central Composite design was successfully used to optimize the parameters of a process that was used for the production of bioethanol by a yeast strain (ZANI *et al.*, 2019). Other studies indicated that CCR designs were also effectively applied in the evaluation of performance and optimization of the parameters of an anaerobic co-digestion of leachate and glycerol for renewable energy generation, and in the maximization of the phytoremediation process of an arsenic-contaminated water (DE SOUZA *et al.*, 2019; TAKEDA *et al.*, 2020). Statistical CCR designs were, as well, able to optimize the parameters for the production of an halotolerant enzyme by a filamentous fungi grown under solid-state fermentation, and the key factors that were affecting the hydrogen production from coffee waste (DAS NEVES *et al.*, 2020; VILLA MONTOYA *et al.*, 2020). The CCR design was also successfully partitioned into blocks without losing the characteristics of rotationality and orthogonality in fertilization experiments, a highly desirable characteristic (CONAGIN, 1982). According to the researchers, CCR designs were capable to study different factors simultaneously, and effectively quantified their individual and combined effects. The mathematical functions constructed by the CCR designs were seen to have a good fit in comparison to the experimental data, and efficiently enabled the synthesis of response surface and contour plots. The authors stated that CCR statistical designs could be successfully used to control, forecast, and optimize scientific experiments and industrial processes.

$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 +$$

$$\beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + e \qquad (30)$$

where $y$ refers to the predicted response of the adjusted design, $\beta_0$ to the value of the intercept, $x_1$, $x_2$ and $x_3$ to the predictor variables, $\beta...$ to the partial linear coefficients, and $e$ to the adjustment error.
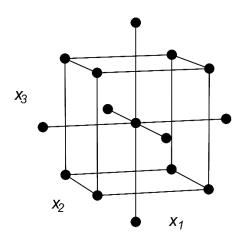
Figure 3 - Representation of a Central Composite Rotatable (CCR) statistical design with two levels, three factors and six-star points ($2^3$, $\alpha = 8^{1/4} = 1,6818$).

## 21 Asymmetrical designs

In some applications, due to the lack of uniform conditions, such as heterogeneous experimental units, restrictions of experimental procedures, factors that have more than two levels, or to the addition of qualitative variables, such as a control group, it has not been possible to include all factor combinations in the factorial designs (MASON *et al.*, 2003; SHESKIN, 2011; MONTGOMERY, 2017). In those situations, however, it has been possible to use factorial arrays that have an unequal number of levels of factors, so called Asymmetrical Factorial (AF) ones (HINKELMANN and KEMPTHORNE, 2007). The designs, also named Mixed-Level factorial designs, have been defined as factorial experiments that have more than two groups of factors with different numbers of levels, in which, all factors in the same group have the same number of levels. For example, $2^m$ x $3^n$ experiments, where the *m* factor has 2 levels each and *n* factor 3 levels each (HINKELMANN and KEMPTHORNE, 2007). AF designs have been mainly evaluated by employing a factorial analysis of variance for a mixed design, where one of the factors is analyzed as a between-subjects variable and the other factor as a within-subjects variable (SHESKIN, 2011). The method has been used to extract the variability as a separate sum of squares, which reduces the magnitude of the error, enables the simultaneously evaluation of the influence of the treatments, and recognize the presence of carryover effects (SHESKIN, 2011), that is, when the effects of the treatments are not independent from each other.

AF designs, according to NAZIEF *et al.* (2020) and KUMAR *et al.* (2017), were successfully applied to evaluate the nutritional and cooking characteristics of brown rice at different storage structure, and to optimize the formulation of solid lipid nanoparticles that

enhances the oral bioavailability of poorly water-soluble drugs. They were also effectively used to analyze the performance of a heat exchanger system, optimize the spherical agglomeration crystallization method of a pharmaceutical active agent, and to investigate the empirical result that an audience experienced to repetition of anthropomorphic ads on multiple-media conditions (GYULAI *et al.*, 2018; AGRAWAL *et al.*, 2020; CHANDRASEKARAN *et al.*, 2021). According to the researchers, AF designs, besides highly flexible due to their capacity to evaluate different type of variables with mixed number of levels, were also capable to economically identify important factors that were significantly contributing to the system's responses, quantify the main effects and interactions, as much to optimize the operational parameters in order to obtain the best desired conditions.

## 22 Sequential experimentation

A comprehensive investigation, which is generally performed to answer questions, frequently involves a sequence of experiments (MASON *et al.*, 2003). The experimental process, therefore, has been used to provide the knowledge to answer the stated questions and, mostly, foment decisions about further experimentations and extend the investigation (MASON *et al.*, 2003; HINKELMANN and KEMPTHORNE, 2007). The first step of a sequence of experiments has conventionally been a screening experiment, used to identify the key variables that will be examined more comprehensively in subsequent experiments, also termed dominant factors (MASON *et al.*, 2003; RODRIGUES and IEMMA, 2005). Screening designs, defined as a deliberately confounded multifactor experiment that has the objective to filter out important main effects, in other words, to "screen out" factors that presents major impact on the dependent variable, have been developed to efficiently evaluate a large number of factors employing a minimal number of observations and limited experimental resources (SHESKIN, 2011; MONTGOMERY, 2017).

The Plackett-Burman (PB) designs (PLACKETT and BURMAN, 1946), e.g., are Fractional Factorial designs that have been used to screen out a large number of factors using the smallest possible number of combinations of experimental conditions. This type of design, where each independent variable has two levels, allows the testing of all the main effects but none of their interactions (SHESKIN, 2011). After the screening process, more extensive investigations involving only the dominant factors have been typically conducted (MASON *et al.*, 2003). The strategy has mainly involved the sequential use of FF designs, which provides the identification of strong two-factor interactions, followed by the use of response surface methods, such as CF or CCR designs, which enables the comprehensive evaluation of few factors, the derivation of second order models and the geometric representation of the responses (MASON *et al.*, 2003). If the objective of the response surface is to determine specific operational regions or optimal conditions, another sequential procedure is often conducted in order to improve the system until the point of desired response (MONTGOMERY, 2017), a procedure named "steepest ascent" (BOX and WILSON, 1951).

The use of PB designs as a screening experiment, followed by the sequential use of Complete, Central Composite or Central Composite Rotational designs, e.g., were successfully employed to develop and evaluate the properties of an oil-based emulsion gel, optimize the production of an enzyme by an *actinomycete* in submerged fermentation, and

to develop a method of sterols and squalene extraction and determination in cyanobacteria (CÂMARA *et al.*, 2020; FAGUNDES *et al.*, 2021; PATEL *et al.*, 2021). The initial use of FF designs, followed by the sequential use of CF and/or Central Composite designs, in turn, was used by Keijok *et al.* (2019) to optimize a method used in the synthesis of metallic nanoparticles, by Gautério *et al.* (2020) to maximize the production of an enzyme by a yeast-like fungus using a by-product of rice grain milling, and by Lv *et al.* (2019) to optimize an ethanol-water distillation column. According to the authors, the results obtained in the initial screening designs were essential to, in an economical number of experiments, identify significant factors that were affecting the system's response. The sequential use CF, Central Composite and CCR designs, indeed, enabled the researchers to effectively evaluate the effects of the selected variables, their interactions, and also to apply response surface methods to optimize the system's parameters in order to obtain ideal specifications.

## 23 Cost analyses

The use of factorial designs has been gaining increasingly importance in costs analyses, also referred as economic evaluations. Those designs have mainly focused in the development of cost-efficient processes and products by using particular criteria, like the expenditures related to specific raw materials, machinery settings, technology employed, fabrication or assembly methods (ASKIN and GOLDBERG, 1988; TACK and VANDEBROEK, 2004). Along that, cost related statistical designs have also aimed at meeting all the product or process functional requirements, obtain the highest possible quality, as well to reduce the environmental negative impacts (SIVAKUMAR *et al.*, 2008; LABIDI *et al.*, 2021). The Taguchi design (TAGUCHI, 1985), e.g, is a Fractional Factorial that was developed to aid the selection of various parameter settings using a minimum number of experimental conditions, and to identify combinations of levels which produce highest robustness, quality performance and low variability over uncontrollable factors (SHESKIN, 2011). The design, firstly developed to improve the quality of manufactured products, has now been frequently applied to design processes to operate consistently and optimally over a variety of conditions, thus, maximize efficiencies and minimize costs (KARNA and SAHAI, 2012).

Taguchi designs, according to reports, were effectively applied to evaluate the economic aspects of the operational conditions of advanced oxidation pre-treatment processes of a reverse osmosis concentrate, the mechanical properties and associated costs of pervious concrete mixtures, as well to analyze and optimize the design of a cost-effective planar waveguide solar concentrator (CAI *et al.*, 2020; KANT *et al.*, 2021; TAHERI and RAMEZANIANPOUR, 2021). Others designs, like the FF, CF, Central Composite and Mixed ones, were used to evaluate economic and environmental aspects of a continuous electrocoagulation process of nitrate removal, analyze the economic sustainability of a methanol production plant using renewable energy sources, and optimize the formulation of cleaning products with better properties and cost (OCHOA *et al.*, 2017; BELLOTTI *et al.*, 2019; KARAMATI-NIARAGH *et al.*, 2019). The authors stated that the factorial designs, in a minimum number of experiments, were capable to identify significant variables that were affecting the system's quality and economic value, estimate optimal parameters conditions, and successfully develop cost-effective products and processes.

## 24 Time-Series studies

Factorial designs have also been increasingly employed in time-series studies, regarded as designs with longitudinal data or repeated measures, as they present measurements repeated over a specific length of time (HINKELMANN and KEMPTHORNE, 2007). Time-series studies, also known as Time-trends, have been often applied in medical, pharmaceutical, nutritional, agricultural or psychological applications, where it is intended to study the influence of the treatments over a certain period of time (HINKELMANN and KEMPTHORNE, 2007). Moreover, they have been used in the evaluation of industrial processes due to their natural dynamic behavior and unstable statistical control (VANHATALO *et al.*, 2013). The key aspect of a time-series design is that the time points have a natural temporal ordering and lack independence, since their errors are correlated (CHRISTENSEN, 1996; KUTNER *et al.*, 2005). In some statistical models, moreover, the time points may also not be equidistant, that is, may be placed in unequally spaced intervals of time (HINKELMANN and KEMPTHORNE, 2007). Time-series factorial designs have been usually analyzed by comparing the time points separately, where each time point is considered as a separate experiment, or jointly through the use of linear filter models, autoregressive models, moving average (ARIMA) and nonstationary models (STEINBERG, 1988; VANHATALO *et al.*, 2013; BOX *et al.*, 2015). The analyses have been also performed based on summarizing measurements, a measure that summary the entire set of time points for each treatment, such as the averages responses for each run (HINKELMANN and KEMPTHORNE, 2007; VANHATALO *et al.*, 2013).

Factorial designs, e.g, were used in a 24-day time-series analysis to investigate the effects and mechanisms that the carbon and nitrogen amendment had on the mineralization of organic phosphorus in microcosm soils, in a 12-month longitudinal study to investigate the role that a pancreatic enzyme supplementation had on patients after a gastric surgery, and in a time-course study related to bone aging gene expression in mice (CATARCI *et al.*, 2018; MISE *et al.*, 2020; WU *et al.*, 2021). They were also effectively used to study daily variations and annual seasonal patterns that fire rotation interval and overstory vegetation type had on ambient soil temperatures in a pine forest, assess the effects of 30 years of nutrition intervention on total and cancer mortality in a population, and to understand and forecast the short and long term of ports demand (WANG *et al.*, 2018; WEISE *et al.*, 2019; DA SILVA and BARBOSA, 2020). According to the researchers, factorial designs, when applied to time-series studies, besides capable to detect experimental time patterns and generate robust diagnoses regarding the dynamics of the systems under study, were also capable to jointly estimate factorial and longitudinal effects, and accurately forecast future events based on previously observed values. Some recent applications of factorial statistical designs in different scientific fields published in the literature were described in table 1.

Table 1 - Applications of factorial statistical designs found in the literature

| Design | Field | Objective | Reference |
|---|---|---|---|
| Complete Factorial | Pharmacology | Optimization of the extraction method of a chemical compound with pharmacological activity from a plant's leaves | Henn *et al.* (2019) |
| Complete Factorial | Environmental | Determination of the optimum efficiency condition of an eco-friendly biosorption process of a cotton dye | Moghazy *et al.* (2019) |
| Complete Factorial | Engineering | Evaluation of the performance of anticorrosive alloys | Porto *et al.* (2019) |
| Complete Factorial | Physics | Investigation and optimization of the dielectric properties of a magneto-dielectric composite | Lara *et al.* (2019) |
| Complete Factorial | Bioprocess | Screening the significant medium constituents that affected the growth rate and final biomass concentration of a freshwater microalga | Ravindran *et al.* (2020) |
| Fractional Factorial | Biotechnology | Development and optimization of a fermentation medium for the production of a biopolymer by a lactic bacterium | Pan *et al.* (2019) |
| Fractional Factorial | Therapeutic | Optimization of drug combinations | Lim *et al.* (2020) |
| Fractional Factorial | Bioscience Engineering | Investigation of the impacts of various constituents on the microbial activity of a plant growth media | Van Gerrewey *et al.* (2020) |
| Fractional Factorial | Civil Engineering | Study the effects that design parameters had on the performance of reinforced concrete bridge piers | Wakjira *et al.* (2020) |

| Design | Field | Objective | Reference |
|---|---|---|---|
| | | strengthened with steel-reinforced polymer composites | |
| Fractional Factorial | Chemistry | Produce cost effective and efficient magnetic activated carbon for the adsorptive removal of pharmaceuticals from aqueous media | Rocha *et al.* (2021) |
| Fractional Factorial | Marketing | Select the best marketing strategy for the purpose of increasing the sales revenue | Al-Dawalibi *et al.* (2020) |
| Central Composite | Process Engineering | Optimization of the parameters of the production process of bioethanol by a *Saccharomyces cerevisiae* strain | Zani *et al.* (2019) |
| Central Composite Rotational | Environmental | Monitoring and maximization of the phytoremediation process of arsenic-contaminated water | De Souza *et al.* (2019) |
| Central Composite Rotational | Environmental | Evaluation of the performance and parameters optimization of an anaerobic co-digestion of leachate and glycerol for renewable energy generation | Takeda *et al.* (2020) |
| Central Composite Rotational | Bioprocess | Optimization of the parameters for the production of an halotolerant enzyme by a filamentous fungi grown under solid-state fermentation | Das Neves *et al.* (2020) |
| Central Composite Rotational | Chemical Engineering | Optimization of the fermentation conditions that were affecting the hydrogen production from coffee waste | Villa Montoya *et al.* (2020) |
| Central Composite Rotational | Agricultural | Field fertilization experiments | Conagin (1982) |
| Asymmetrical | Pharmaceutics | Study the effects and optimize the formulation of solid lipid nanoparticles that enhances | Nazief *et al.* (2020) |

| Design | Field | Objective | Reference |
|---|---|---|---|
| | | the oral bioavailability of poorly water-soluble drugs | |
| Asymmetrical | Food Science | Evaluate the nutritional and cooking characteristics of brown rice at different storage structures | Kumar *et al.* (2017) |
| Asymmetrical | Engineering | Analyze the performance of a heat exchanger system | Chandrasekaran *et al.* (2021) |
| Asymmetrical | Pharmaceutical | Optimize the spherical agglomeration crystallization method of an active agent | Gyulai *et al.* (2018) |
| Asymmetrical | Marketing | Investigate the effects that an audience experienced to anthropomorphic ads on multiple-media condition | Agrawal *et al.* (2020) |
| Plackett-Burman and Complete | Food Technology | Development and evaluation of the properties of an oil-based emulsion gel | Câmara *et al.* (2020) |
| Plackett-Burman and Central Composite | Biotechnology | Optimization of the production of an enzyme by an *actinomycete* | Patel *et al.* (2021) |
| Plackett-Burman and Central Composite Rotational | Chemistry | Develop a method of sterols and squalene extraction and determination in cyanobacteria | Fagundes *et al.* (2021) |
| Fractional and Complete | Biotechnology | Optimization of a method used for the biosynthesis of metallic nanoparticles | Keijok *et al.* (2019) |
| Fractional and Central Composite | Biotechnology | Maximization of the production of an enzyme by a yeast-like fungus using a by-product of rice grain milling | Gautério *et al.* (2020) |
| Fractional, Complete and Central | Engineering | Optimization of an ethanol-water distillation column | Lv *et al.* (2019) |

| Design | Field | Objective | Reference |
|---|---|---|---|
| Composite Face-Centered | | | |
| Central Composite and Central Composite Rotational | Biotechnology | Optimization of the enzyme production by a fungus isolated from the Atlantic Forest | Vieira *et al.* (2019) |
| Taguchi | Chemical Engineering | Evaluate the economic aspects of the operational conditions of advanced oxidation pre-treatment processes of a reverse osmosis concentrate | Cai *et al.* (2020) |
| Taguchi | Mechanical Engineering | Analyze and optimize the design of a cost-effective planar waveguide solar concentrator | Kant *et al.* (2021) |
| Taguchi | Civil Engineering | Evaluate the mechanical properties and associated costs of pervious concrete mixtures | Taheri & Ramezanianpour (2021) |
| Complete and Central Composite | Engineering | Analyze the economic sustainability of a methanol production plant using renewable energy sources | Bellotti *et al.* (2019) |
| Fractional and Mixture | Chemistry | Optimization of the formulation of cleaning products with better properties and cost | Ochoa *et al.* (2017) |
| Central Composite | Environmental Engineering | Evaluate economical and environmental aspects of a continuous electrocoagulation process of nitrate removal | Karamati-Niaragh *et al.* (2019) |
| Complete Factorial | Agricultural | 24-day time-series analysis to investigate the effects and mechanisms that the carbon and nitrogen amendment had on the mineralization of organic phosphorus in soil | Mise *et al.* (2020) |
| Complete Factorial | Clinical | 12-month longitudinal study to investigate the role that a | Catarci *et al.* (2018) |

| Design | Field | Objective | Reference |
|--------|-------|-----------|-----------|
| | | pancreatic enzyme supplementation had on patients after a gastric surgery | |
| Complete Factorial | Biomedical | Time-course study related to bone aging gene expression in mice | Wu *et al.* (2021) |
| Complete Factorial | Ecology | Study daily and annual patterns that fire rotation interval and overstory vegetation had on ambient soil temperatures in a pine forest | Weise *et al.* (2019) |
| Fractional Factorial | Medical | Assess the effects of 30 years of nutrition intervention on total and cancer mortality in a population | Wang *et al.* (2018) |

## Conclusions

Through the use of matrices with all the possible treatment combinations, factorial statistical designs have been extremely useful in scientific experiments and research activities. They have been used to draw accurate inferences from the experimental results, estimate confidence and prediction intervals, characterize relationships among the variables, quantify the individual and combined effects of factors, distinguish assignable causes of variations from random ones, as well to estimate the adequacy and validate the proposed models. The fractionalization of factorial designs has been successfully employed to considerably reduce the number of experiments without the loss of pertinent information. The property has been useful to screen out the variables of interest in experiments with a large number of factors. The orthogonality and rotationality (rotatability) characteristics were achieved by the addition of star (alpha) and central points to the designs' matrices. The condition has provided the achievement of robust estimates regarding the individual and combined effects of the variables, fit responses with curvatures, and derive quadratic polynomial functions regarding the nature of the responses.

From reports found in the literature, it was observed that the Complete, Fractional, Central Composite Rotational and Asymmetrical factorial statistical designs were successfully applied in investigations related to different scientific fields. The designs were used to economically identify variables that were significantly affecting the responses of the systems under study, quantify the magnitude of experimental errors, as well the individual and combined influence of the factors. They were also used to derive mathematical functions that accurately represented the behavior of the responses, elaborate response surface and contour plots, and optimize the system's conditions in order to obtain desired specifications. Moreover, the designs were capable to, in a minimum number of

experiments, develop cost-effective products and processes within specific quality standards, diagnose the dynamics of longitudinal investigations, and accurately forecast future events based on previously observed values. Due to their high efficiency and robustness, statistical factorial designs will most likely be keep increasingly used in the future. As addressed in this article, their great applicability, which allow them to be employed in many academic and industrial areas, make them a powerful and important tool in modern researches and development of technology.

## Acknowledgements

- *RESUMO: Planejamentos fatoriais têm sido cada vez mais utilizados em investigações científicas e no desenvolvimento tecnológico. Os delineamentos, por meio da utilização de matrizes com todas as combinações de tratamentos, têm sido capazes de caracterizar efetivamente as relações entre as variáveis de experimentos multifatoriais, avaliar as variabilidades experimentais e derivar funções matemáticas que representam o comportamento das respostas. Os delineamentos fatoriais foram fracionados, o que reduziu substancialmente o número de tratamentos sem a perda de informações relevantes. A adição de pontos centrais e estrela às matrizes fatoriais conferiu a eles as características de ortogonalidade e rotatividade, frequentemente utilizadas para ajustar modelos com curvatura e identificar regiões críticas de interesse. De acordo com os relatos da literatura os delineamentos fatoriais, também chamados de experimentos fatoriais, foram aplicados com sucesso em diferentes tipos de investigações, incluindo avaliações de custos e estudos de séries temporais. Eles foram capazes de estimar características importantes dos experimentos tais como efeitos individuais e combinados dos fatores, a magnitude dos resíduos, além de expressar as relações das variáveis em equações polinomiais, desenhar gráficos de superfície e contorno de resposta e determinar combinações ótimas de parâmetros. Nesta revisão, os aspectos fundamentais dos planejamentos Fatoriais Completo, Fracionado, Rotacional Composto Central e Assimétrico foram apresentados e as aplicações recentes dessas ferramentas poderosas foram descritas.*

- *PALAVRAS-CHAVE: Delineamento de experimentos; DOE; planejamento experimental; otimização de processo; desenho experimental; planejamento estatístico; modelo estatístico.*

## References

AGRAWAL, S.; KHANDELWAL, U.; BAJPAI, N. Anthropomorphism in advertising: the effect of media on audience attitude. *Journal of Marketing Communications*, p.1–17, 2020.

AL-DAWALIBI, A.; AL-DALI, I. H.; ALKHAYYAL, B. A. Best marketing strategy selection using fractional factorial design with analytic hierarchy process. *MethodsX*, v.7, p.100927, 2020.

ALTMAN, D. G.; BLAND, J. M. Standard deviations and standard errors. *BMJ*, v.331, n.7521, p.903, 2005.

ANDERSON, M. J.; WHITCOMB, P. J. Design of Experiments. *In*: *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, 2010.

ANGELOPOULOS, P.; EVANGELARAS, H.; KOUKOUVINOS, C. Run orders for efficient two level experimental plans with minimum factor level changes robust to time trends. *Journal of Statistical Planning and Inference*, v.139, n.10, p.3718–3724, 2009.

ASKIN, R. G.; GOLDBERG, J. B. Economic Optimization in Product Design. *Engineering Optimization*, v.14, n.2, p.139–152, 1988.

BELLOTTI, D.; CASSETTARI, L.; MOSCA, M.; MAGISTRI, L. RSM approach for stochastic sensitivity analysis of the economic sustainability of a methanol production plant using renewable energy sources. *Journal of Cleaner Production*, v.240, p.117947, 2019.

BEZERRA, M. A.; SANTELLI, R. E.; OLIVEIRA, E. P.; VILLAR, L. S.; ESCALEIRA, L. A. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, v.76, n.5, p.965–977, 2008.

BHOWMIK, A.; VARGHESE, E.; JAGGI, S.; VARGHESE, C. Minimally changed run sequences in factorial experiments. *Communications in Statistics - Theory and Methods*, v.46, n.15, p.7444–7459, 2017.

BOX, G. E. P.; HUNTER, J. S. Multi-Factor Experimental Designs for Exploring Response Surfaces. *Annals of Mathematical Statistics*, v.28, n.1, p.195–241, 1957.

BOX, G. E. P.; HUNTER, J. S. The $2^{k-p}$ Fractional Factorial Designs. *Technometrics*, v.3, n.3, p.311–351, 1961.

BOX, G. E. P.; WILSON, K. B. On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, v.13, n.1, p.1–45, 1951.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

BROWNLEE, K. A. *Statistical theory and methodology in science and engineering*. Wiley, 1965.

CAI, Q. Q.; WU, M. Y.; LI, R.; DENG, S. H.; LEE, B. C. Y.; ONG, S. L.; HU, J. Y. Potential of combined advanced oxidation – Biological process for cost-effective organic matters removal in reverse osmosis concentrate produced from industrial wastewater reclamation: Screening of AOP pre-treatment technologies. *Chemical Engineering Journal*, v.389, p.123419, 2020.

CÂMARA, A. K. F. I.; OKURO, P.K.; SANTOS, M.; PAGLARINI, C. de S.; DA CUNHA, R. L.; RUIZ-CAPILLAS, C.; HERRERO, A. M.; POLLONIO, M. A. R. Understanding the role of chia (Salvia Hispanica L.) mucilage on olive oil-based emulsion gels as a new fat substitute in emulsified meat products. *European Food Research and Technology,* v.246, n.5, p.909–922, 2020.

CATARCI, M.; BERLANDA, M.; GRASSI, G. B.; MASEDU, F.; GUADAGNI, S. Pancreatic enzyme supplementation after gastrectomy for gastric cancer: a randomized controlled trial. *Gastric Cancer*, v.21, n.3, p.542–551, 2018.

CHANDRASEKARAN, R.; QIAN, X.; LEE, S. Thermal performance evaluation and analysis of the efficient and sustainability shell and tube heat exchanger system. *Journal of Construction Project Management and Innovation*, v.11, n.1, p.151–159, 2021.

CHENG, C.-S.; MARTIN, R. J.; TANG, B. Two-Level Factorial Designs with Extreme Numbers of Level Changes. *The Annals of Statistics*, v.26, n.4, p.1522–1539, 1998.

CHRISTENSEN, R. *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. CRC Press, 1996.

CONAGIN, A. Delineamentos "compostos centrais ortogonais, rotacionais e divisíveis em blocos". *Bragantia*, v.41, n.1, p.49–56, 1982.

COOK, R. D. Detection of Influential Observation in Linear Regression. *Technometrics*, v.19, n.1, p.15–18, 1977.

COX, D. R.; HINKLEY, D. V. *Theoretical Statistics*. CRC Press, 1979.

D'AGOSTINO, R. B. *Goodness-of-Fit-Techniques*. CRC Press, 1986.

DA SILVA, F. G. F.; BARBOSA, R. B. *Applying Factorial Models to Forecast Ports Demands*. SSRN Scholarly Paper, n.ID 3520285. Rochester, NY: Social Science Research Network, 2020.

DAS NEVES, C. A.; DE MENEZES, L. H. S.; SOARES, G. A.; DOS SANTOS REIS, N.; TAVARES, I. M. C.; FRANCO, M.; DE OLIVEIRA, J. R. Production and biochemical characterization of halotolerant β-glucosidase by Penicillium roqueforti ATCC 10110 grown in forage palm under solid-state fermentation. *Biomass Conversion and Biorefinery*, 2020.

DE SOUZA, T. D.; BORGES, A. C.; BRAGA, A. F.; VELOSO, R. W.; TEIXEIRA DE MATOS, A. Phytoremediation of arsenic-contaminated water by Lemna Valdiviana: An optimization study. *Chemosphere*, v.234, p.402–408, 2019.

DRAPER, N.R.; SMITH, H. *Applied Regression Analysis*. John Wiley & Sons, 1998.

FAGUNDES, M. B.; ALVAREZ-RIVERA, G.; VENDRUSCOLO, R. G.; VOSS, M.; DA SILVA, P.A.; BARIN, J. S.; JACOB-LOPES, E.; ZEPKA, L. Q.; WAGNER, R. Green microsaponification-based method for gas chromatography determination of sterol and squalene in cyanobacterial biomass. *Talanta*, v.224, p.121793, 2021.

FISHER, R. A. Design of Experiments. *British Medical Journal*, v.1, n.3923, p.554–554, 1936.

FISHER, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, v.52, n.2, p.399–433, 1919.

GALTON, F. *Natural Inheritance*. Macmillan and Company, 1889.

GAUTÉRIO, G. V.; DA SILVA, L. G. G.; HÜBNER, T.; DA ROSA RIBEIRO, T.; KALIL, S. J. Maximization of xylanase production by Aureobasidium pullulans using a by-product

of rice grain milling as xylan source. *Biocatalysis and Agricultural Biotechnology*, v.23, p.101511, 2020.

GRUBBS, F. E. Sample Criteria for Testing Outlying Observations. *Annals of Mathematical Statistics*, v.21, n.1, p.27–58, 1950.

GUNST, R. F.; MASON, R. L. Fractional factorial design. *WIREs Computational Statistics*, v.1, n.2, p.234–244, 2009.

GYULAI, O.; KOVÁCS, A.; SOVÁNY, T.; CSÓKA, I.; AIGNER, Z. Optimization of the Critical Parameters of the Spherical Agglomeration Crystallization Method by the Application of the Quality by Design Approach. *Materials*, v.11, n.4, p.635, 2018.

HENN, J. G.; STEFFENS, L.; DE MOURA SPEROTTO, N.D.; DE SOUZA PONCE, B.; VERÍSSIMO, R. M.; BOARETTO, F. B. M.; HASSEMER, G.; PÉRES, V.F.; SCHIRMER, H.; PICADA, J. N.; SAFFI, J.; MOURA, D. J. Toxicological evaluation of a standardized hydroethanolic extract from leaves of Plantago australis and its major compound, verbascoside. *Journal of Ethnopharmacology*, v.229, p.145–156, 2019.

HILOW, H. Comparison among run order algorithms for sequential factorial experiments. *Computational Statistics & Data Analysis*, v.58, p.397–406, 2013.

HINKELMANN, K.; KEMPTHORNE, O. *Design and Analysis of Experiments* (Vol. 1). New York: Wiley, 2007.

HOAGLIN, D. C.; MOSTELLER, F.; TUKEY, J. W. *Fundamentals of Exploratory Analysis of Variance.* John Wiley & Sons, 1991.

HOOPER, D.; COUGHLAN, J.; MULLEN, M. R. Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, v.6, n.1, p.53–60, 2008.

KALINOWSKI, P.; FIDLER, F. Interpreting Significance: The Differences Between Statistical Significance, Effect Size, and Practical Importance. *Newborn and Infant Nursing Reviews*, v.10, n.1, p.50–54, 2010.

KANT, K.; NITHYANANDAM, K.; PITCHUMANI, R. Analysis and Optimization of a Novel Hexagonal Waveguide Concentrator for Solar Thermal Applications. *Energies*, v.14, n.8, p.2146, 2021.

KARAMATI-NIARAGH, E.; ALAVI MOGHADDAM, M. R.; EMAMJOMEH, M. M.; NAZLABADI, E. Evaluation of direct and alternating current on nitrate removal using a continuous electrocoagulation process: Economical and environmental approaches through RSM. *Journal of Environmental Management*, v.230, p.245–254, 2019.

KARNA, S. K.; SAHAI, R. An overview on Taguchi method. *International journal of engineering and mathematical sciences*, v.1, n.1, p.11-18, 2012.

KEIJOK, W. J.; PEREIRA, R. H. A.; ALVAREZ, L. A. C.; PRADO, A. R.; DA SILVA, A. R.; RIBEIRO, J.; DE OLIVEIRA, J. P.; GUIMARÃES, M. C. C. Controlled biosynthesis of gold nanoparticles with Coffea arabica using factorial design. *Scientific Reports*, v.9, n.1, p.16019, 2019.

KUMAR, A.; BHOPLE, S.; KUMAR, N.; TIWARI, V.K. Effect of Ageing, Moisture Contents and Storage Structures on Nutritional and cooking characteristics of brown rice during storage. *Chemical Science Review and Letters*, v.6, n.22, p.793-800, 2017.

KUTNER, M. H.; NACHTSHEIM, C. J.; NETER, J.; LI, W. *Applied Linear Statistical Models*. McGraw-Hill Irwin, 2005.

LABIDI, A.; TANABE, I.; TAKAHASHI, S. A study on extending technologies lifespan for the environment safety. *Journal of Machine Engineering*, p.109–120, 2021.

LARA, L. A.; MANCIPE, D. L.; PINEDA, Y.; MORENO, J. J.; PEÑA-RODRÍGUEZ, G. Design and characterization of a magneto-dielectric composite in high frequency with aligned magnetite powders. *Journal of Physics: Conference Series*, v.1386, p.012103, 2019.

LIM, J. J.; GOH, J.; RASHID, M. B. M. A.; CHOW, E. K.-H. Maximizing Efficiency of Artificial Intelligence-Driven Drug Combination Optimization through Minimal Resolution Experimental Design. *Advanced Therapeutics*, v.3, n.4, p.1900122, 2020.

LV, S.; HE, Z.; QUEVEDO, A. V.; MIRABILE, Y. Z.; VINING, G. G. Process optimization using sequential design of experiment: A case study. *Quality Engineering*, v.31, n.3, p.473–483, 2019.

MASON, R. L.; GUNST, R. F.; HESS, J. L. *Statistical Design and Analysis of Experiments: With Applications to Engineering and Science*. John Wiley & Sons, 2003.

MILES, J.; SHEVLIN, M. A time and a place for incremental fit indices. *Personality and Individual Differences*, Special issue on Structural Equation Modeling. v.42, n.5, p.869–874, 2007.

MISE, K.; MARUYAMA, R.; MIYABARA, Y.; KUNITO, T.; SENOO, K.; OTSUKA, S. Time-series analysis of phosphorus-depleted microbial communities in carbon/nitrogen-amended soils. *Applied Soil Ecology*, v.145, p.103346, 2020.

MOGHAZY, R. M.; LABENA, A.; HUSIEN, Sh. Eco-friendly complementary biosorption process of methylene blue using micro-sized dried biosorbents of two macro-algal species (Ulva fasciata and Sargassum dentifolium): Full factorial design, equilibrium, and kinetic studies. *International Journal of Biological Macromolecules*, v.134, p.330–343, 2019.

MONTGOMERY, D. C. *Design and Analysis of Experiments*. John Wiley & Sons, 2017.

MORETTIN, P.A.; BUSSAB, W. O. *Estatística Básica.* Saraiva Educação S.A., 2017.

MULAIK, S. A.; JAMES, L. R.; VAN ALSTINE, J.; BENNETT, N.; LIND, S.; STILWELL, C. D. Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, v.105, n.3, p.430–445, 1989.

NAKAGAWA, S.; CUTHILL, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, v.82, n.4, p.591–605, 2007.

NAKAGAWA, S.; SCHIELZETH, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, v.4, n.2, p.133–142, 2013.

NAZIEF, A. M.; HASSAAN, P.S.; KHALIFA, H. M.; SOKAR, M. S.; EL-KAMEL, A. H. Lipid-Based Gliclazide Nanoparticles for Treatment of Diabetes: Formulation,

Pharmacokinetics, Pharmacodynamics and Subacute Toxicity Study. *International Journal of Nanomedicine*, v.15, p.1129–1148, 2020.

NOORDIN, M. Y.; VENKATESH, V.C.; SHARIF, S.; ELTING, S.; ABDULLAH, A. Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *Journal of Materials Processing Technology*, v.145, n.1, p.46–58, 2004.

OCHOA, D. H.; BRAGA, J. W. B.; MACHADO, F. Optimization of the Functional Characteristics of Cleaning Products Through Experimental Design. *Journal of Surfactants and Detergents*, v.20, n.2, p.467–481, 2017.

ONYIAH, L. C. *Design and Analysis of Experiments: Classical and Regression Approaches with SAS*. CRC Press, 2008.

PAN, N.C.; BIZ, G.; BALDO, C.; CELLIGOI, M. A. P.C. Factorial design in fermentation medium development for hyaluronic acid production by Streptococcus zooepidemicus. *Acta Scientiarum. Technology*, v.42, p.e42729–e42729, 2019.

PATEL, G. B.; SHAH, K. R.; SHINDHAL, T.; RAKHOLIYA, P.; VARJANI, S. Process parameter studies by central composite design of response surface methodology for lipase activity of newly obtained Actinomycete. *Environmental Technology & Innovation*, v.23, p.101724, 2021.

PATINO, C. M.; FERREIRA, J. C. Confidence intervals: a useful statistical tool to estimate effect sizes in the real world. *Jornal Brasileiro de Pneumologia*, v.41, n.6, p.565–566, 2015.

PEARSON, K. Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London*, v.185, p.71–110, 1894.

PLACKETT, R. L.; BURMAN, J. P. The design of optimum multifactorial experiments. *Biometrika*, v.33, n.4, p.305–325, 1946.

PORTER, S. C.; VERSEPUT, R. P.; CUNNINGHAM, C. R. Process optimization using design of experiments. *Pharmaceutical technology*, v.21, n.10, p.60-71, 1997.

PORTO, M. B.; PORTELA, D. G.; NETO, A. F. de A. Temperature, current density and cobalt concentration effects on electrodeposited anticorrosive cobalt-tungsten alloys using factorial experiment design and ANOVA techniques. *Transactions of the IMF*, v.97, n.6, p.305–311, 2019.

RAVINDRAN, B.; KASSIM, M. A.; MOHAMED, M. S. Screening of Medium Constituents for the Cultivation of *Scenedesmus dimorphus* UTEX 1237 using $2^k$ Factorial Design Approach. *IOP Conference Series: Materials Science and Engineering,* v.716, p.012003, 2020.

ROCHA, L. S.; SOUSA, É. M. L.; GIL, M. V.; OLIVEIRA, J. A. B. P.; OTERO, M.; ESTEVES, V.I.; CALISTO, V. Producing Magnetic Nanocomposites from Paper Sludge for the Adsorptive Removal of Pharmaceuticals from Water—A Fractional Factorial Design. *Nanomaterials*, v.11, n.2, p.287, 2021.

RODRIGUES, M. I.; IEMMA, A. F. *Planejamento de Experimentos e Otimização de Processos: Uma Estratégia Sequencial de Planejamentos.* Casa do Pão Editora, 2005.

SHESKIN, D. J. *Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition.* CRC Press, 2011.

SIVAKUMAR, K.; SARAVANAN, R.; NOORUL HAQ, A. Cost-tolerance modelling and optimisation of machining tolerance design through intelligent techniques. *International Journal of Machining and Machinability of Materials*, v.3, n.1–2, p.162–189, 2008.

STEIGER, J. H. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, v.25, n.2, p.173–180, 1990.

STEINBERG, D. M. Factorial Experiments With Time Trends. *Technometrics*, v.30, n.3, p.259–269, 1988.

TACK, L.; VANDEBROEK, M. Trend-resistant and cost-efficient cross-over designs for mixed models. *Computational Statistics & Data Analysis*, v.46, n.4, p.721–746, 2004.

TAGUCHI, G. Quality engineering in Japan. *Communications in Statistics - Theory and Methods*, v.14, n.11, p.2785–2801, 1985.

TAHERI, B. M.; RAMEZANIANPOUR, A. M. Optimizing the mix design of pervious concrete based on properties and unit cost. *Advances in concrete construction*, v.11, n.4, p.285-298, 2021.

TAKEDA, P.Y.; GOTARDO, J. T.; GOMES, S. D. Anaerobic co-digestion of leachate and glycerol for renewable energy generation. *Environmental Technology*, p.1–11, 2020.

TSAI, C.-L.; CAI, Z.; WU, X. The Examination of Residual Plots. *Statistica Sinica*, v.8, n.2, p.445–465, 1998.

TUKEY, J. W. The Future of Data Analysis. *The Annals of Mathematical Statistics*, v.33, n.1, p.1–67, 1962.

VAN GERREWEY, T.; AMELOOT, N.; NAVARRETE, O.; VANDECRUYS, M.; PERNEEL, M.; BOON, N.; GEELEN, D. Microbial activity in peat-reduced plant growing media: Identifying influential growing medium constituents and physicochemical properties using fractional factorial design of experiments. *Journal of Cleaner Production,* v.256, p.120323, 2020.

VANHATALO, E.; BERGQUIST, B.; VÄNNMAN, K. Towards Improved Analysis Methods for Two-Level Factorial Experiments with Time Series Responses: Analysis Methods for Experiments with Time Series Responses. *Quality and Reliability Engineering International*, v.29, n.5, p.725–741, 2013.

VIEIRA, F. G. N.; CHRIST, D.; GRACIANO, L.; CORRÊA, J. M.; KADOWAKI, M. K.; SILVA, J. L. da C.; GANDRA, R. F.; MALLER, A.; POLI, M. de L. T. de M.; SIMÃO, R. de C. G. Experimental Design for Optimization of β-Xylosidase Production by A. fumigatus Isolated from the Atlantic Forest (Brazil). *Journal of Advances in Biology & Biotechnology*, p.1–16, 2019.

VILLA MONTOYA, A. C.; DA SILVA MAZARELI, R. C.; DELFORNO, T. P.; CENTURION, V.B.; DE OLIVEIRA, V.M.; SILVA, E. L.; VARESCHE, M. B. A. Optimization of key factors affecting hydrogen production from coffee waste using factorial design and metagenomic analysis of the microbial community. *International Journal of Hydrogen Energy*, v.45, n.7, p.4205–4222, 2020.

WAKJIRA, T. G.; NEHDI, M. L.; EBEAD, U. Fractional factorial design model for seismic performance of RC bridge piers retrofitted with steel-reinforced polymer composites. *Engineering Structures*, v.221, p.111100, 2020.

WANG, S.-M.; TAYLOR, P.R.; FAN, J.-H.; PFEIFFER, R. M.; GAIL, M. H.; LIANG, H.; MURPHY, G. A.; DAWSEY, S. M.; QIAO, Y.-L.; ABNET, C. C. Effects of Nutrition Intervention on Total and Cancer Mortality: 25-Year Post-trial Follow-up of the 5.25-Year Linxian Nutrition Intervention Trial. *JNCI: Journal of the National Cancer Institute*, v.110, n.11, p.1229–1238, 2018.

WEISE, D. R.; SACKETT, S. S.; HAASE, S. M.; JOHNSON, N.Effects of fire rotation interval and overstory type on ambient soil temperatures in ponderosa pine forests in Arizona. *Canadian Journal of Forest Research*, v.49, n.10, p.1320–1328, 2019.

WEISSMAN, S. A.; ANDERSON, N.G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Organic Process Research & Development*, v.19, n.11, p.1605–1633, 2015.

WU, J.; GUPTA, M.; HUSSEIN, A. I.; GERSTENFELD, L. Bayesian modeling of factorial time-course data with applications to a bone aging gene expression study. *Journal of Applied Statistics*, v.48, n.10, p.1730–1754, 2021.

YULE, G. U. On the Theory of Correlation. *Journal of the Royal Statistical Society*, v.60, n.4, p.812–854, 1897.

ZANCAN, G. T. Educação científica: uma prioridade nacional. *São Paulo em Perspectiva*, v.14, n.3, p.3–7, 2000.

ZANI, S. H. M.; ASRI, F. M.; AZMI, N.S.; YUSSOF, H. W.; ZAHARI, M. A. K. M. Optimization of process parameters for bioethanol production from oil palm frond juice by *Saccharomyces cerevisiae* using response surface methodology as a tool. *IOP Conference Series: Materials Science and Engineering,* v.702, p.012003, 2019.

ZWILLINGER, D.; KOKOSKA, S. *CRC Standard Probability and Statistics Tables and Formulae*. CRC Press, 1999.