Research Article

© 2024 Soto-Rodríguez et al.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/)

Received: 20 December 2023 / Accepted: 26 February 2024 / Published: 5 March 2024

Survey on the Compliance of the Assumptions of Variance Analysis of Experimental Designs in Undergraduate Theses of Peruvian Universities

Iván Dennys Soto-Rodríguez

Frida Rosa Coaquira Nina

Jorge Chue Gallardo

Fernando René Rosas Villena

Rino Nicanor Sotomayor Ruiz

Academic Department of Statistics and Informatics,
Universidad Nacional Agraria La Molina,
Av. la Molina s/n,
La Molina,
Perú

DOI: https://doi.org/10.36941/jesr-2024-0043

Abstract

The main objective of the research was to verify compliance with the assumptions of the analysis of variance, using some experimental design, in undergraduate theses from Peruvian universities. For this purpose, 200 undergraduate theses from professional degrees in Agricultural Sciences from 13 Peruvian universities were chosen at random. A descriptive textual analysis was carried out with the sentiment analysis classification algorithm and the text mining technique was used to discover patterns and trends. The result showed that, in almost all of the theses, the verification of the two most important assumptions of the analysis of variance, normality and homogeneity of variances, were not taken into consideration. Ignoring the importance of assumptions and the significance they have on the validity of research results, forces us to pay attention to the way in which the statistical methodology of experimental designs is being carried out in undergraduate theses.

Keywords: *experimental designs, assumptions, normality, homogeneity of variances*

1. Introduction

In scientific research, experimental designs are the main statistical method used to test the hypothesis that the mean effect of the treatments under study is similar or that the mean effect of one of them is different from the others.

(Kirk, 2009) points out that experimental design is a plan by which treatments are managed under certain specific conditions. He asserts that there are five closely related activities: formulating hypotheses, specifying critical regions for testing them, collecting observations according to the

experimental design employed, analyzing the collected information, and interpreting the conclusions drawn from hypothesis testing. (McGuigan, 1993) indicates that experimental designs possess six important characteristics: recognizing the experimental problem, formulating hypotheses, handling or manipulating the treatments (independent variables), operationalizing the treatment comparison variable (dependent variable), selecting the design for the research, employing the most suitable techniques for data analysis, and specifying the objectives expected from the application of the experimental design. (Montgomery, 2013) mentions that experimental designs are used for comparative studies of processes or systems. He conceptualizes it as a process where the levels of a factor (treatments) transform the input data of the process (experimental units) into output values of the process (observations). (González, 2014) defines the design of experiments as a systematic statistical technique aimed at conducting a series of tests to introduce predetermined changes, with the goal of identifying the effects that treatments produce on the experimental units through a variable under study.

(Fisher, 1960) emphasizes that to correctly apply an experimental design, three fundamental principles must be taken into consideration: the need for sufficient repetitions to ensure accurate estimation of the variance of the experimental error, the randomization of treatments among experimental units (or vice versa) to ensure the independence of observations, and the control of experimental error to facilitate fair comparison of treatments.

In the application of experimental designs, several mistakes can lead to an increase in experimental error. In the initial stage, errors may arise from inadequately choosing the experimental design to be used. During the application process, errors can occur such as those associated with verifying the assumptions of residual (independence, normality, and homogeneity of variances), those related to repetitions, errors in constructing the analysis of variance table, and in the final stage of the process, the error of inadequately interpreting the research findings. The study focuses on errors related to the assumptions of normality and homogeneity of variance of the residual.

Theoretical knowledge of experimental design methodology is essential for researchers to adequately apply experimental designs. This knowledge is typically imparted in university courses on statistical methods for research. However, it's important to recognize that understanding experimental designs goes beyond mere theoretical learning (Ocampo Castaña, 2018). In this sense, (Diaz Levicoy, Aguayo Arriagada, & Cortés Toro, 2014) point out that teaching with projects allows a better understanding and application of the concepts used in experimental designs. This leads to the following research questions: Do students of professional careers linked to agricultural sciences in Peruvian universities verify the assumptions of experimental designs when they prepare their undergraduate thesis? What are the most frequent errors in the application of experimental designs? How do thesis students use the words associated with the fulfillment of assumptions in analysis of variance?

To address these research questions, a random sample of 200 research papers from the agricultural sciences programs of 13 Peruvian national universities, conducted between 2010 and 2020, was taken into account.

Experimental designs are statistical methodologies belonging to the field of parametric techniques. They rely on a specific probability distribution that the treatment comparison variable must adhere to. In the literature review, various definitions of experimental designs are found, for example: (Chapin, 1950) indicates that an experimental design is the observation of results that are not related to the procedures used to exercise control, (Cox & Reid, 2000) point out that the term experiment in a research implies that all procedures must be established by the researcher. On the other hand, (González, 2014) considers that experimental design as a systematic statistical technique aimed at conducting a series of tests to induce predetermined changes, with the goal of identifying treatments that significantly influence the variable of interest. (Montgomery, 2013) mentions that the statistical design of experiments "refers to the process of planning the experiment so that appropriate data are collected and analyzed using statistical methods, which will result in valid and objective conclusions". (Raudonius, 2017) states that papers published in agricultural and biological journals

often do not correctly utilize experimental statistical techniques.

Regarding the assumptions required by experimental designs, they fall into two categories: those that are universal across all experimental designs and those that uniquely characterize each design.

In the universal group, two assumptions are of interest in this research due to their impact on the statistical inference process: the normality of experimental errors and the homogeneity of variability among the samples receiving the treatments.

On the non-compliance of assumptions, (Pelea, 2018) argues that in many investigations experimental designs are used with experimental data that do not comply with the assumptions of normality and homoscedasticity which are required to be able to employ a parametric statistical method.

On the other hand, (Eisenhart, 1947) identifies four assumptions: random variables, additivity, equal variances and zero correlations and normality of the errors. (Kuehl, 2001) maintains that the validity of the estimations and the verification of the hypotheses associated with the linear additive model are conditional on the compliance that the experimental errors must be independent, normally distributed with zero mean and common variance between the treatments being compared. But, (Knief & Forstmeier, 2021) indicate that Gaussian models are robust to non-normality except for data with influential outliers, (Stroup, 2015) says that applying ANOVA to non-normally distributed data leads to a substantial loss of power and significant inaccuracy.

In this sense, (Schielzeth, y otros, 2020) conclude that mixed-effects models are great alternative to severe violations of model assumptions like normality and (Blanca Mena, Alarcón Postigo, Arnau Gras, Bono Cabré, & Bendayan, 2017) indicates that F-test in terms of Type I error was robust for non-normality distributions used in the health and social sciences.

Experimental design models are classified according to three types of effects: fixed, random and mixed. A model is fixed effects when the parameters are considered constant, random effects when the effects are random variables and mixed when some effects are fixed and others are random (Sahai & Ageel, 2000).

These models have been analyzed from various perspectives, including: the effects to be estimated, the response variable under consideration, the treatments to be compared, the identification of experimental units, the methods for randomization and blocking, the optimal sample size, and the appropriate number of replicates. As far as experimental errors are concerned, (Cochran, 1947) indicates that the most certain way to prevent experimental errors from being excessively correlated is the random assignment of the treatments to the experimental units or vice versa, as well as the random selection of the samples of experimental units that are going to receive the treatments under study.

In this research, the interest is for models are fixed by the researcher (fixed effects) because according to (Bell & Jones, 2015) they are frequently used in economics, political science and agricultural science. According to (Bell, Fairbrother, & Jones, 2019) these models do not provide parameter estimates when the assumption of error normality is not met. Authors such as (Hill, Davis, Roos, & French, 2020) recommend not using fixed effects techniques when the effects of variables change over time. (Eisenhart, 1947) points out that when applying analysis of variance to models that have fixed effects, the following assumptions must be taken into consideration for the conclusions to be valid: the observed values are random variables with mean mij, the effects are additive, the variances are equal and the covariances are equal to zero, and that the observed values are jointly distributed as a multivariate normal.

On the other hand, (Fernández, Trapero, & Domínguez, 2010) indicate that the results of the analysis of variance in fixed effects models will be valid if the residuals are independent and have a normal behavior with uniform variance and if the factors that are part of the model have additive effects. Also, they point out that ANVA is a "robust" technique; that is, its results change very little due to non-compliance with some of the assumptions mentioned above. (Tabachnick, 2007) indicates that when applying an experimental design the following aspects should be considered:

independence of errors, different sample sizes, normality of sample distributions, extreme values, homogeneity of variances, effect of sample size, parameter estimation, post results comparisons and analysis of non-compliance with homogeneity of variances if necessary.

(Curtis, et al., 2018) point out that common errors when using experimental designs are the lack of randomization and blinding, the use of different group sizes and the application of statistical analysis to samples smaller than 5. Also, they indicate that to apply ANVA, the assumptions of normality and homogeneity of variances must be verified first. In relation to normality, (Curtis, et al., 2018) mention that if the data do not have normal distribution, techniques from the field of parametric statistics such as ANVA should not be used. In the face of this drawback, nonparametric techniques such as the Mann-Whitney test can be used.

In the study by (Harwell, Rubinstein, Hayes, & Olds, 1992) to study the robustness of the F test in one- and two-factor experimental designs, it was found that the type I error was sensitive to non-compliance with homogeneous variances. In relation to this, (Lindman, 2012) suggests that the aforementioned assumptions are rarely met with real data, highlighting the impact of non-normality and heterogeneity on the appropriateness of the F test for the null hypothesis, particularly noting that nonzero kurtosis significantly affects the F test.

The scope of the research is restricted to the case of undergraduate theses of thirteen Peruvian national universities in which experimental design models with fixed effects are used and in which the assumptions of normality and homogeneity of variances should be verified. This study is based on the academic concern of how these designs are used in professional theses.

2. Material and Methods

The research is quantitative, non-experimental and cross-sectional of the theses collected in the year 2022, taken at random from the thesis repositories of Peruvian public universities. The research variables are application errors in experimental designs; The aspects to be considered for the evaluation have to do with errors present in the methodology applied throughout the development of the thesis, the application of the appropriate experimental design and the evaluation of the assumptions.

The population comprises all theses from agricultural sciences faculties of the 22 national universities. A random sample of 200 theses from 13 national universities offering agricultural sciences specialties was selected. Sample determination utilized a 5% risk and a maximum error of 6.95%, assuming p=q=0.5. The sampling method employed two-stage sampling, selecting 13 universities offering agricultural sciences specialties, and randomly choosing theses from each university to complete the sample size of 200.

Descriptive, exploratory, and multivariate techniques were used, along with text mining to specify textual result trends. Initially, researchers perceived a lack of understanding of concepts such as homogeneity of variances, normality of errors, and independence of observations of the response variable among thesis authors. This intuition of the lack of understanding of statistical terms pointed out by (Ocampo Castaña, 2018) and (Diaz Levicoy, Aguayo Arriagada, & Cortés Toro, 2014) led to the decision to carry out a text mining analysis with the tm library of the R software (R Core Team, 2023) and a text association analysis using the Jaccard coefficient (Polo Bautista & Martínez Acevedo, 2021), of the materials and methods chapters, and results and discussion of the 200 theses considered in the aleatory sample. The Jaccard index is defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Where:

A = Data set 1

B= Data set 2

|A∩B| = Number of elements in common in both sets

|AUB| = Number of unique elements from both sets

3. Results

The application of text mining to the theses was conducted with a minimum frequency threshold of 600 to filter out words that do not significantly contribute to identifying the appropriate use of statistics. Figure 1 depicts the resulting word cloud. This cloud reveals those words such as "experimental," "average," "variance," "Duncan," "difference," and "blocks," among others, are utilized by the authors of the theses, aligning with the topic of experimental designs. Additionally, the proportion of the use of these words is at least 600/200 = 3 times per thesis.

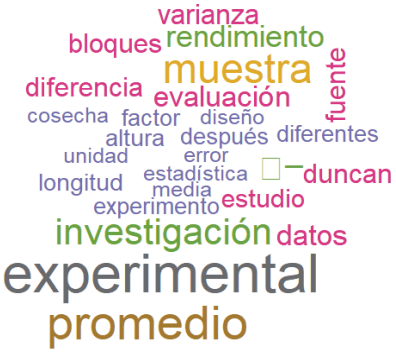


Figure 1. Word cloud of the methodology, results and conclusions chapters.
Source: own elaboration based on 200 undergraduate theses from national universities

Table 1 shows the percentage distribution of theses that use a specific experimental design.

Table 1. Percentage distribution of experimental designs. Source: self made.

Experimental Design	Percentage
DBCA with split plots	5.70
Complete Randomized Block Design	46.49
Complete Randomized Design	25.00
Factorial experiment in a CRBD	12.32
Others	10.49
Total	100.00

Table 2 shows the percentage distribution of the theses that verified the assumption of normality of errors.

Table 2. Percentage distribution of the number of theses that verified the normality of the errors.

Is the normality of the errors verified?	Percentage
No	90.35
There is no clear information	3.07
Yes	6.58
Total	100.00

Source: Self made

Table 3 presents the percentage distribution of the theses that verified the homogeneity of variances of the treatments.

Table 3. Percentage distribution of the theses that verified the homogeneity of variances.

Was the homogeneity of variances of the treatments verified?	Percentage
N	92.54
There is no clear information	2.63
Yes	4.82
Total	100.00

Source: Self made

Table 4 presents the percentage distribution of the theses that verified the assumption of independence of errors.

Table 4. Percentage distribution of theses that verified the test of independence of errors.

Is the test for independence of errors performed?	Percentage
Yes	0.00
No	99.12
There is no clear information	0.88
Total	100.00

Source: Self made

Table 5 presents the percentage distribution of the theses that determined the sample size subjectively.

Table 5. Percentage distribution of theses that determined the sample size subjectively.

¿Was the sample size determined subjectively?	Percentage
Yes	98.25
No	1.75
Total	100

Source: Self made

4. Discussion

From Table 1, it's evident that a significant majority, 71.49%, utilize basic designs such as complete randomized design and complete randomized block design. However, it's crucial to note that compliance with the assumptions inherent to the chosen experimental design must be verified before application.

Table 2 reveals that the normality test of errors was neglected in 90.35% of the theses, despite its paramount importance in parametric statistics.

In Table 3, it's apparent that the assumption was disregarded in 92.54% of the theses, despite its significance in facilitating equitable treatment comparisons.

Table 4 highlights that this assumption wasn't verified in 99.12% of the theses. This test bears similar importance to tests of normality and homogeneity of variances within experimental design analysis.

Table 5 illustrates that in 98.25% of the reviewed theses, the sample size (number of repetitions) was subjectively considered, lacking full justification as required.

Regarding the associations between the words calculated with the Jaccard index, a strong correlation of 0.85 was found between "significant" and "difference," a low correlation of 0.40 between "normality" and "homogeneity," a moderate correlation of 0.64 between "homogeneity" and "variances," and no correlation of the word "independence" with any other word in the 200 theses analyzed. The expression "significant difference" is commonly used when rejecting the null hypothesis, which states that the effects of the treatments are equal. This explains the strong association of 0.85 between these words ("difference" and "significant"). Regarding the poor association between the words "normality" and "homogeneity" of only 0.40, this may be due to the assumption that "y_1, y_2, ..., y_n and n are normal with mean $\mu_i = \mu + \tau_i$ and with variance σ^2 constant," which is not considered in most theses. This is evidenced by the almost total non-existence of the test of homogeneity of variances in 92.54% of the theses, as seen in Table 3. In the same sense, the words "homogeneity" and "variances" have an association of 0.64, which can be interpreted as weak, but it is not tangibly manifested in the theses with the respective proof, apparently being a colloquial expression.

(Lindman, 2012) points out that failure to comply with this assumption has consequences in test F. It is therefore important to ask the following question, what percentage of the theses that do not consider this assumption arrive at correct or incorrect results? The researchers cannot answer this question because the original data (observations) were not found in the theses.

It is necessary to mention that the tests for homogeneity of variances and the test for normality of errors constitute the most important assumptions for the parametric analysis of variance.

5. Conclusions

The use of experimental designs in research necessitates adherence to two fundamental principles: (a) comprehending the respective experimental design regarding its advantages, disadvantages, and limitations, and (b) adhering to the assumptions proposed for the theoretical development of the experimental design to ensure the validity and reliability of the obtained results. The analysis of a random sample of 200 theses from various public universities in Peru, considering the aforementioned principles, resulted in the following conclusions

1. Theses prominently feature terms related to experimental designs; however, there is a low association between the words "Normality" and "Kolmogorov-Smirnov," indicating the limited application of the Kolmogorov-Smirnov test and, therefore, a lack of importance given to the normality of errors. Regarding the concept of "independence," there is a complete absence of association with other words, indicating a limited understanding of this assumption in the 200 analyzed theses.
2. The widespread omission of the normality test of errors in 90.35% of the theses suggests a lack of rigor in evaluating basic statistical assumptions. This compromises the validity of the applied experimental designs, as detailed in Table 1, which could affect the reliability of the obtained results. Additionally, the failure to conduct the homogeneity of variances test in 92.54% of the theses is a concerning finding. The neglect of this test, along with the omission of the normality test of errors, highlights the non-compliance with essential assumptions for parametric analysis of variance. These assumptions are crucial to ensure the validity of statistical inferences made in the study.
3. The astonishing figure of 99.12% of theses not meeting the assumption of independence of errors is a matter of great concern. This finding underscores a widespread lack of attention to an equally crucial assumption. The independence of errors is essential to ensure the validity of applied statistical tests, and its negligence could compromise the proper interpretation of results.
4. The subjective choice of sample size in 98.25% of cases without adequate justification based

on the availability of experimental material is a critical aspect. This suggests a lack of scientific foundation in determining the sample size, potentially affecting the validity and generalization of results. The need to implement more robust and justifiable methods for determining sample size in future research is emphasized.

References

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data*. *Political Science Research and Methods, 3*(1), 133-153.

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random efects models: making an informed choice. *Quality & quantity, 53*, 1051-1074.

Blanca Mena, M., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557.

Chapin, F. S. (1950). Experimental Design in Sociology-Limitations and Abuses. *Soc. F.*, 25.

Cochran, W. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics, 3*(1), 22-38.

Cox, D., & Reid, N. (2000). *The theory of the design of experiments.* CRC Press.

Curtis, M. J., Alexander, S., Cirino, G., Docherty, J. R., George, C. H., Giembycz, M. A., . . . Ahluwalia, A. (2018). Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers. *British Journal of Pharmacology*, 987-993.

Diaz Levicoy, D. A., Aguayo Arriagada, C. G., & Cortés Toro, C. (2014). Enseñanza de la Estadística mediante proyectos y su relación con teorías de aprendizaje. *Premisa*, 16-23.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics, 3*(1).

Fernández, R., Trapero, A., & Domínguez, J. (2010). *Experimentación Agraria.* Córdoba: Ediciones Díaz de Santos.

Fisher, R. (1960). *The design of experiments* (7th ed.). Oliver and Boyd. London and Edinburgh.

González, I. J. (2014). Diseño de experimentos y su aplicación en la industria. *Ingenio y Conciencia Boletín Científico de la Escuela Superior de Cd. Sahagún, 1*(1).

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlos Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics, 17*(4), 315-339.

Hill, T., Davis, A., Roos, J., & French, M. (2020). Limitations of fixed-effects models for panel data. *Sociological Perspectives, 63*(3), 357-369.

Kirk, R. (2009). Experimental design. *Sage handbook of quantitative methods in psychology*, 23-45.

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods, 53*(6), 2576-2590.

Kuehl, R. O. (2001). *Diseño de experimentos. Principios estadísticos de diseño.* México: Thomson Learning.

Lindman, H. (2012). *Analysis of variance in experimental design.* Springer Science & Business Media.

McGuigan, F. (1993). *Experimental psychology: Methods of research.* Prentice-Hall, Inc.

Montgomery, D. (2013). *Design and Analysis.* Arizona: John Wiley & Sons, Inc.

Ocampo Castaña, A. (2018). La comprensión de conceptos estadísticos en la educación secundaria. *Scientia et technica*.

Pelea, L. P. (2018). ¿ Cómo proceder ante el incumplimiento de las premisas de los métodos paramétricos? o¿ cómo trabajar con variables biológicas no normales? *Revista del Jardín Botánico Nacional, 39*, 1-12.

Polo Bautista, L. R., & Martínez Acevedo, K. V. (2021). Algoritmo para el análisis temático de documentos digitales. *Investigación bibliotecológica, 35*(89).

R Core Team. (30 de Noviembre de 2023). *_R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing*. Obtenido de https://www.r-project.org/

Raudonius, S. (2017). Application of statistics in plant and crop research: important issues.

Sahai, H., & Ageel, M. (2000). *The Analysis of Variance: Fixed, Random and Mixed Models.* Boston: Springer Science+Business Media, LLC.

Schielzeth, H., Dingemanse, N., Nakagawa, S., Westneat, D., Allegue, H., Teplitsky, C., . . . Araya-Ajoy, Y. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in ecology and evolution, 11*(9), 1141-1152.

Stroup, W. (2015). Rethinking the analysis of non-normal data in plant and soil science. *Agronomy journal, 107(2), 811-827, 107*(2), 811-827.

Tabachnick, B. G. (2007). *Experimental designs using ANOVA.* Belmont, CA: Thomson/Brooks/Cole.

Zemdirbyste-Agriculture, 104(4).