Revista Ciencia en Desarrollo. Vol. 2 Num. 15. pp. 102 - 106
julio - diciembre 2024

https://doi.org/10.19053/uptc.01217488.v15.n2.2024.17181

Artículo de revisión

Teaching One-Way analysis of variance using country map puzzles

Enseñando análisis de varianza de un solo factor usando rompecabezas del mapa de un país

Simón Cuartas-Rendón[1], Freddy Hernández-Barajas [2], Olga Usuga-Manco[3], Víctor Ignacio López-Ríos[4]

[1]Estadístico, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia.
[2]Ingeniero Industrial, PhD en Estadística, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia.
[3]Ingeniera Industrial, PhD en Estadística, Universidad de Antioquia Medellín, Colombia.
[4]Matemático, PhD en Estadística, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia.

Recepción: 06-feb-2024 **Aceptado:** 15-jun-2024 **Publicado:** 23-jul-2024
Cómo citar: Hernández-Barajas, F., Cuartas-Rendón, S., Usuga-Manco, O., & López-Ríos, V. I. (2024). Enseñando análisis de varianza de un solo factor usando rompecabezas del mapa de un país. Ciencia En Desarrollo, 15(2). https://doi.org/10.19053/uptc.01217488.v15.n2.2024.17181

Abstract

The teaching of single-factor experimental design is crucial for future professionals. Typically, instructors deliver lectures and assign exercises without exploring more engaging didactic methodologies. This research introduces a novel approach: students learn by playing with a puzzle of a country map. They must assemble one of two puzzles: one with department names and one without names. The variable of interest is the time it takes to complete the puzzle. Additionally, a Shiny application is proposed for data logging and analysis. In an experiment involving 25 students, evidence was found that completion time is shorter when the puzzle includes department names. In summary, this innovative methodology aids students in understanding experimental design in the classroom.

Keywords: Experimental design, analysis of variance, active learning.

Resumen

La enseñanza del diseño de experimentos de un solo factor es crucial para futuros profesionales. Por lo general, los profesores imparten clases magistrales y asignan ejercicios, sin explorar metodologías didácticas más cautivadoras. Esta investigación presenta una metodología novedosa: los estudiantes aprenden jugando con un rompecabezas del mapa de un país. Deben armar uno de dos rompecabezas: uno con nombres de departamentos y otro sin los nombres. La variable de interés es el tiempo que tardan en completar el rompecabezas. Además, se propone una aplicación Shiny para registrar datos y análisis. En un experimento con 25 estudiantes, se encontró evidencia de que el tiempo es menor cuando el rompecabezas incluye nombres de departamentos. En resumen, esta metodología innovadora ayuda a los estudiantes a comprender el diseño de experimentos en el aula.

Palabras Clave: Diseño de experimentos, análisis de varianza, aprendizaje activo.

E-ISSN: 2462-7658

1 Introduction

ANOVA models enable the assessment of whether the levels or effects of one or more independent variables significantly differ concerning a response variable. These models constitute core statistical principles studied by students in disciplines such as Statistics, Mathematics, other sciences, and Engineering. To instruct undergraduates on ANOVA models, a variety of pedagogical approaches can be employed. These strategies aim to bolster students' comprehension and facilitate a profound understanding, steering away from rote memorization of statistical concepts associated with ANOVA models.

Johnson [3] explores intuitive approaches utilizing graphical representations and real-life instances to aid students in grasping the rationale and underlying assumptions behind analysis of variance (ANOVA) models. Additionally, within his article, he presents a hypothetical example involving quantities, offering students the opportunity to comprehend how conclusions drawn from an adjusted ANOVA model could vary under conditions of high and low variance.

Lesser and Melgoza [4] explore intuitive approaches utilizing graphical representations and real-life instances to aid students in grasping the rationale and underlying assumptions behind analysis of variance (ANOVA) models. Additionally, within their article, they presents a hypothetical example involving quantities, offering students the opportunity to comprehend how conclusions drawn from an adjusted ANOVA model could vary under conditions of high and low variance.

Campillo-Ferrer et al. [1] discuss the importance of *gamification* during the learning-teaching process for undergraduate students, pointing out that the use of elements or ludical activities related to games have a significantly positive effect on the participation, commintment, and learning of students even when they face challenging problems or new more advanced topics.

Mangiero et al. [5] consider the use of *Microsoft Office Excel* to promote statistical and data science competences on undergraduate and graduate students for many different disciplines, including statistics itself. These authors focus the use of this application to teach probabilistic concepts, such as the Poisson and the normal distributions, confidence intervals, and highlight the non-dependence on internet connection to make this educative strategy possible.

Medina-Hernández et al. [6] reflect through a meta-analysis about the pedagogic strategies that have being implemented in recent decades to teach statistics and data analytics, pointing out the need to alienate students' experiences at the classroon with their expectations and interests through investigation projects that allow them to discuss in the classroom the contents and generate significative knowledge from them.

Riquelme [8] creates an application that allow stundets to learn basic concepts on statistical inference, such us some probabilistic distributions, hypothesis testing or confidence intervals. This app has a clean and intuitive interface, which gives the opportunity to stundets to interact with different values or set of data, so they can understand easily this topics.

Zhao et al. [9] study how an R application can make learning process of concepts related to the normal distribution or Cohen's *d* easier. Moreover, they found that even though there is no significant difference on the learning of these concepts when this app is not used, the students highlighted that having this kind of tools available make it more interesting for them to review this concepts after class.

This article is structured as follows: In Section 2, we explain the experimental design with a single factor. In Section 3, we introduce the proposed experiment for data collection, and in Section 4, we analyze the collected data. Finally, in Section 6 we present some conclusions and recommendations for instructors interested in replicating the experiment with their students.

2 Experimental design with a single factor

The experimental design of only one factor is a statistical methodology that allows to determine if the effect of the a levels or treatments of one independent variable over on variable of interest are significantly equal or not. In this way, when one is interested on the equality of more than one means, analysis of variance —also known as ANOVA— is greatly useful. To fit thi kind of models, n repetitions of the experiment for each of the a treatments are made.

According to Montgomery [7], if Y_{ij} is the value taken by the variable of interest (also known as dependent variable or response variable), when the independent variable takes its i-th treatment at the j-th repetition of the experiment. Therefore, the analysis of variance for the experimental design of one factor might be expressed as follows:

$$\begin{aligned} Y_{ij} &= \mu + \tau_i + \varepsilon_{ij}, \\ \varepsilon_{ij} &\sim NID(0,\, \sigma^2), \\ i &= 1,\, \ldots,\, a, \\ j &= 1,\, \ldots, n, \end{aligned} \tag{1}$$

where μ is the global mean, meaning that this a common parameter for every treatment, τ_i is a parameter associated exclusively to the i-th treatment, and ε_{ij} is the random error related to every repetition for each treatment. Note that errors are assumed independet and identically distributed following a normal distribution with null mean and constant variance (this implies that the variance is independent of both a repetition and treatment).

For the performance of the experiment, once the amount of n repetitions is known for every single treatment (when this quantity is constant for every treatment we talk about a balanced experiment), the experimenter must randomize the execution of each repetition, so the inclusion of new fonts of variance that affect the assumptions over the errors is avoided.

The information gotten during the performance of the experiment is usually registred as shown in Table 1:

Table 1: Table to register the information in an experiment with a single factor.

Treatment (i)	Observations				Total	Mean
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$\frac{y_{a.}}{y_{..}}$	$\frac{\bar{y}_{a.}}{\bar{y}_{..}}$

By other hand, we talk about a fixed effects ANOVA model when levels are chosen arbitrarily by the experimenter. In contrast, when the levels included in the experiment are chosen from a random sample of a universe of levels, we are studying a random effects models. For this paper, we will consider a fixed effects model, for which it is assumed that the sum of the effects associated to each treatment is zero, so we can have unique estimators for each parameter through ordinary least squared (they are the same estimators found with maximum likelihood).

That said, in order to determine if the treatments have effects, on average, different over the response variable, we can prove the following hypothesis:

$$H_0: \tau_i = 0 \; \forall \, i \quad vs. \quad H_1: \tau_i \neq 0, \; \exists \, i$$

where, given a significance level α, the test statistic is:

$$F = \frac{MS_{Treatments}}{MS_{Error}} = \frac{SS_{Treatments}/(a-1)}{SS_{Error}/(na-1)}, \tag{2}$$

where $F \sim F_{a-1,\, na-1}$. Moreover, if F_c is the calculated statistic, the the p-value (V_p) can be calculated just as follows:

$$V_p = \Pr[F_{a-1,\, na-1} > F_c] \tag{3}$$

3 Proposed experiment

In this section, we explain an experiment created to ensure that graduated students in an Experimental Design course understand and comprehend the concepts. Additionally, the goal of the experiment is that each student generates data that has personal significance for them, as they are required to reach out to volunteer students to conduct the experiment.

With this experiment, we aim to determine whether there is a significant difference in the average time it takes for an individual to complete a political-administrative country map puzzle, depending on whether the map puzzle includes names or not. Figure 1 shows the country map puzzle, on the left a map with names to identify the states easily, and on the right the same map without names.

Figure 1: Maps used in the workshop. On the left a map without names, on the right a map with names.

In this experiment, the independent variable is the time that is needed to solve the map puzzle, putting every single piece in its right place. This means that there are two levels: $i = 1$ when the map puzzles have names, and $i = 2$ when the map puzzle does not have any names. The variable Y_{ij} is the time that the j-th person needs to solve the type i map puzzle, this magnitude can be obtained in minutes and seconds and then be transformed into seconds only. The volunteers are randomly assigned to each level (with names or without names) using a coin toss. This ensures the fulfillment of a crucial assumption in the experimental design with a single factor.

Each individual in the experiment must participate only once with one type of map. This must be adhered to because if two observations are taken for each map type from the same individual, there's a high likelihood of introducing bias during the second attempt, as the person will already be familiar with the map puzzle's figures.

To facilitate the execution of the experiment, we have created a workshop in Microsoft Word. This workshop includes the activity's objectives, the required materials, guidelines for approaching volunteers, tables for data collection, and a set of questions to answer using the collected data. The workshop is available in two languages, allowing instructors to download, customize, and utilize it with their students. Below are links to the workshops in English and Spanish.

- Workshop in English: visit the URL.
- Workshop in Spanish: visit the URL.

Instructors could use the workshop in the classroom to perform the experiment and to obtain meaningful data. Figure 2 shows a photo of two real participants doing the experiment in the classroom. At the top we observe a participant making a map puzzle with names and below a participant with a puzzle without names.

Figure 2: Images capturing the ongoing experiment display distinct scenarios. The top photo exhibits the map puzzle featuring names, whereas the bottom one portrays the map puzzle devoid of any labels.

4 Application of the experiment and Results

This section presents the dataset and analysis after applying the proposed experiment with 25 university students at Universidad Nacional de Colombia and Universidad de Antioquia in the first semester of 2023. In the following table, we have the dataset collected. The first column corresponds to a counter, the second represents the time (in minutes) to complete the map puzzle correctly and the third column contains the information about the type of the map puzzle.

The initial step in the analysis involves a graphical examination of the response variable. Figure 3 presents a boxplot illustrating the time required to complete the map puzzle for different map puzzle types (with and without names). From this figure, we can observe that the boxplot for maps without names tends to have higher values compared to the boxplot for maps with names. Both boxplots exhibit

Table 2: Dataset collected after the application of the experiment.

i	Time (min)	Type of map puzzle
1	4.83	With names
2	18.13	Without names
3	18.50	With names
4	28.23	Without names
5	25.77	Without names
6	16.80	With names
7	16.75	Without names
8	10.95	With names
9	8.88	With names
10	28.30	Without names
11	11.43	With names
12	12.72	Without names
13	15.03	With names
14	15.45	Without names
15	20.27	Without names
16	13.97	With names
17	17.20	Without names
18	18.03	With names
19	9.72	With names
20	20.70	With names
21	15.53	Without names
22	14.65	With names
23	21.88	Without names
24	8.85	Without names
25	13.12	Without names

similar variability and are nearly symmetrical.

Table 3 presents the results of the ANOVA. The p-value, reported as 0.0298 in the table, suggests that there is a significant effect attributed to the type of map.

Table 3: ANOVA table with the results.

Source	Df	Sum Sq	Mean Sq	F value	p-value
Map	1	156.36	156.36	5.37	0.0298
Res	23	669.97	29.13		

Figure 4 shows the Q-Q plot for the 25 residuals. From this plot we observe that residuals do not appear to be normal, indeed, the Shapiro test generates a p-value of 0.8299, indicating that errors have a normal distribution.

5 Shiny application to analize the data

In this section, we present an interactive html application created with the Shiny R package [2]. This application facilitates the analysis of the experimental dataset as shown in Table 2.

Instructors and students could use the application. By default, the application uses the same dataset shown in Table 2 but any user could upload his own dataset. The user is required to choose both the response variable and the factor. Subsequently, the application will generate a boxplot comparing the time, a QQ plot depicting the residuals, and the ANOVA table. Figure 5 shows a screenshot of the application with inputs and outputs.

The proposed app can be found in the next URL:

https://fhernanb.shinyapps.io/exp_map_puzzle/

The app can also be executed locally by copying the next code in the R console:

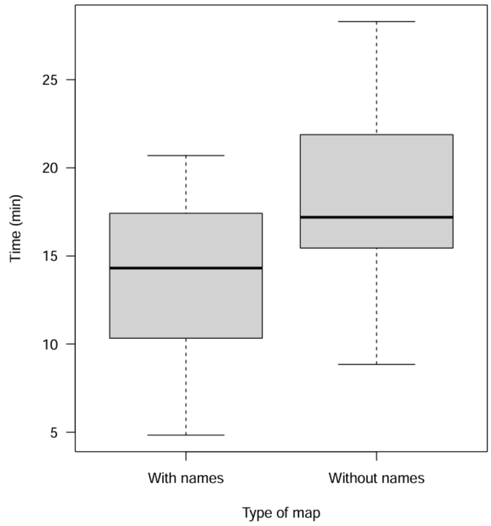

Figure 3: Boxplot for the time to complete the map puzzle given each type of map.

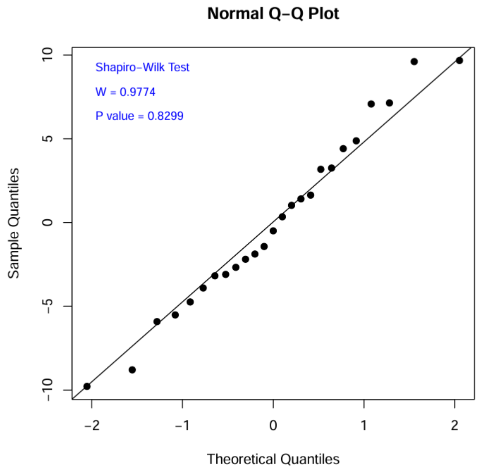

Figure 4: Normal Q-Q plot for the residuals.

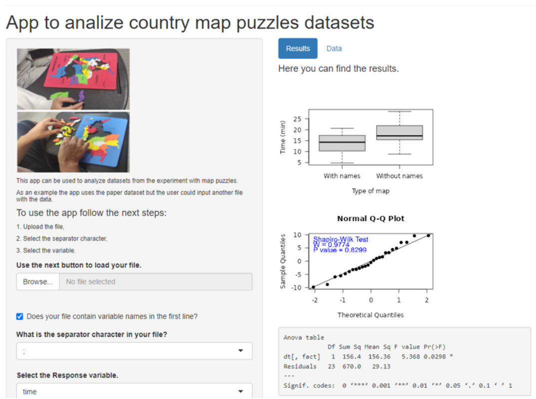

Figure 5: Screenshot of the shiny application to analyze experimental datasets.

```
if (!require("shiny"))
 install.packages("shiny")
library(shiny)
runGitHub(repo="semilleroApps",
          user="fhernanb",
          sub="exp_map_puzzle")
```

6 Conclusions

This article introduces an innovative and straightforward experiment employing a country map puzzle to demonstrate the practical application of ANOVA within a classroom setting. Instructors teaching statistics and experimental design courses can easily replicate this experiment in different countries. The materials needed for this experiment are minimal: two country maps (one labeled with state names and one without) and a stopwatch for time recording purposes.

After conducting the experiment, it became apparent that participants given the map puzzle without state names took a longer time to complete it, while those provided with the map puzzle containing state names finished more quickly. The gathered data underwent statistical analysis, verifying the assumptions of ANOVA. The results strongly suggested significant differences in the time taken to complete the map puzzle.

Figure 6: Example of other map puzzles to replicate the workshop.

References

[1] J.-M. Campillo-Ferrer, P. Miralles-Martínez, and R. Sánchez-Ibáñez. Gamification in higher education: Impact on student motivation and the acquisition of social and civic key competencies. *Sustainability*, 12(12), 2020.

[2] W. Chang, J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, and B. Borges. *shiny: Web Application Framework for R*, 2022. URL https://CRAN.R-project.org/package=shiny. R package version 1.7.3.

[3] D. Johnson. An intuitive approach to teaching analysis of variance. *Teaching of Psychology*, 16(2):67–68, 1989.

[4] L. Lesser and L. Melgoza. Simple numbers: Anova example of facilitating student learning in statistics. *Teaching Statistics*, 29: 102–105, 2007.

[5] G. A. Mangiero, A. Qayyum, and C. J. Cante. Teaching statistics – a dynamic excel approach. *Journal of Education for Business*, 97(7):439–444, 2022. doi: 10.1080/08832323.2021.1974815. URL https://doi.org/10.1080/08832323.2021.1974815.

[6] E. J. Medina-Hernández, J. L. Muñiz, D. S. Guzmán-Aguiar, and A. Holguín-Higuita. Recursos y estrategias para la enseñanza de la estadística y la analítica de datos en la educación superior. *Formación superior universitaria*, 15:61 – 68, 06 2022. ISSN 0718-5006. URL http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-50062022000300061&nrm=iso.

[7] D. Montgomery. *Design and Analysis of Experiments*. Student solutions manual. John Wiley & Sons, 2008. ISBN 9780470128664. URL http://books.google.de/books?id=kMMJAm5bD34C.

[8] V. Riquelme. Enseñanza de la estadística inferencial mediante una aplicación móvil. *Revista Latinoamericana de Investigación en Matemática Educativa*, 23, 05 2021. URL https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-24362020000200233#B13.

[9] F. Zhao, L. Schützler, O. Christ, and R. Gaschler. Learning statistics with interactive pictures using r shiny: Generally preferred, but not generally advantageous. *Teaching Statistics*, 45 (2):106–124, 2023. doi: https://doi.org/10.1111/test.12324. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/test.12324.