Available online at http://www.journalijdr.com

ISSN: 2230-9926

IJDR

International Journal of Development Research
Vol. 12, Issue, 06, pp. 56652-56655, June, 2022
https://doi.org/10.37118/ijdr.24647.06.2022

RESEARCH ARTICLE OPEN ACCESS

THE VAN DER WAERDEN TEST AND WELCH'S ANOVA: A SIMULATED STUDY

Igor Campos de Almeida Lima*[1], Marcello Montillo Provenza[1,2], Paulo Henrique Couto Simões[1], Antônio Alexandre Lima[2] and Jorge Luiz de Jesus Goulart[1]

[1]Statistics Department, State University of Rio de Janeiro, Rio de Janeiro, Brazil
[2]Mathematics Department, State University of Rio de Janeiro, São Gonçalo, Brazil

ARTICLE INFO

Article History:
Received 17th March, 2022
Received in revised form
03rd April, 2022
Accepted 28th May, 2022
Published online 22nd June, 2022

Key Words:
Kruskall-Wallis test, Welch's ANOVA, Van der Waerden test, R Software.

Corresponding author:
Igor Campos de Almeida Lima

ABSTRACT

In the undergraduate Brazilian courses, the approach to experimental statistics is reduced in the menu so that professors have time to explain all the content. However, some methodologies are not used, such as Welch's ANOVA and the van der Waerden test. The objective of this work is to present these two techniques that are neglected, however, can contribute to more robust analyzes in other experiments. Using the *R* program four treatment variables were simulated and the proposed tests were applied. The results showed that these two methods can be used as good alternative solutions when certain assumptions are violated. Thus, these two procedures could be mostly explored by university Brazilian professors.

Citation: ***Igor Campos de Almeida Lima, Marcello Montillo Provenza, Paulo Henrique Couto Simões, Antônio Alexandre Lima, and Jorge Luiz de Jesus Goulart.*** "The van der waerden test and welch's anova: a simulated study*", International Journal of Development Research*, 12, (06), 56652-56655.

INTRODUCTION

In different undergraduate courses such as Engineering, Economics, Administration, and others, the content of applied statistics is taught concisely, in order to meet the academic calendar of the Brazilian institutions. Typically, the content is divided into descriptive statistics and inference. Within statistical inference, there are parametric tests (those whose distribution parameter is known) and non-parametric tests. The parametric test used when comparing multiple groups is the Analysis of Variance (ANOVA), with its non-parametric equivalent being the Kruskal-Wallis test. The ANOVA is a method of comparing means. The method is used when you want to test the null hypothesis that the means of k groups have equal means (with k more than 2). The alternative hypothesis is that at least one group has a mean different from the others. Among the required assumptions, data must be collected randomly, with Normal distribution, present constant variance (homoscedasticity), and independence (LEVINE, et al. 2016; FONTELLES, 2012). The Kruskal-Wallis test is the non-parametric equivalent of the one-way ANOVA (LEVINE, et al., 2016; VIERA, 2010; DAWSON & TRAPP 2003; SIEGEL, 1979). In both the single-factor ANOVA and the Kruskal-Wallis test, if the null hypothesis is rejected, one must proceed to two new phases: one to verify the assumptions and the other to apply post-hoc tests to identify which groups differ. Different textbooks like STOCKEMER (2019), NAGADEVARA (2019), LEVINE, et al. (2016), HEIBERGER & HOLLAND (2015), DAALGARD (2008), BRANDT (2014), WEISBURD & BRITT (2014), FONTELLES (2012), BLACK (2019), WALPOLE, et al. (2009), TRIOLA (2008), MENDENHALL (1985), WONNACOTT & WONNACOTT (1980), and SPIEGEL (1978) cover the content of ANOVA with some addressing the Kruskal-Wallis test. However, among the books cited, some indicate that if there is a violation of the assumption of normality, the Kruskal-Wallis test should be used, but they do not indicate what to do if the assumption of homoscedasticity is violated.

Welch's ANOVA can be used when the assumption of homoscedasticity is violated (ZAIONTZ, 2020). The van der Waerden test has the advantages of both the one-way ANOVA (when the assumptions are not violated) and the Kruskal-Wallis test (when the ANOVA assumptions are violated) (NIST, 2015; SPRENT & SMEETON, 2001). Therefore, this work aims to use two tests not addressed at the undergraduate level: the van der Waerden test and Welch's ANOVA.

MATERIAL AND METHODS

The set of information analyzed in this article consists of a matrix with four variables which can also be called treatments (*Trat1*, *Trat2*, *Trat3*, and *Trat4*). The values were simulated in a random sample of size 12. The data are presented in Table 1. The simulation and analysis were performed in the R software, version 3.6.1, with the *stats*, *ggplot2*, and *e1071* packages (R CORE TEAM, 2019). In the ANOVA one way, the total variation of the data is subdivided into two parts: one referring to the variations that are attributed to differences between the groups and another part referring to the oscillations that are attributed to differences within groups (HEIBERGER & HOLLAND, 2015). The variability between groups is obtained from the square sum between the groups while the variation within the groups is obtained from the square sum within the groups (LEVINE, et al., 2016; FONTELLES, 2012; TRIOLA, 2008). The ratio between the quadratic sum and the degree of freedom provides the mean square, which is used to calculate the F test, at a given significance level α. After building the ANOVA, if the null hypothesis is rejected, one should investigate which pairs of groups differ from each other and any possible violation of the assumptions. There are different tests in the literature to verify the distribution of the data, such as Shapiro-Wilk, Jarque-Bera, D'Agostino-Pearson, Shapiro-Francia, and others. In addition, you can use some graphic resources such as the QQ-plot. There are several tests to assess the homoscedasticity of data, such as the Bartlett, Cochran, Hartley, and Brown-Forsythe tests (LEVINE et al., 2016; HEIBERGER & HOLLAND, 2015; ZAR, 2014). The Brown-Forsythe test is a robust measure for the presence of outliers, as it uses the median to obtain the absolute differences for each new value. This new data matrix is used to calculate ANOVA with the same degrees of freedom in the numerator and denominator, both in the F test and in obtaining the critical value. Welch's ANOVA can be used after the homoscedasticity test indicates that the study data are heteroscedastic, especially when the samples have different sizes (ZAIONTZ, 2020). The construction is based on the following equation:

$$= \frac{\frac{1}{c-1}\sum_{j=1}^{c}\left(\frac{n_j}{s_j^2}\right)\left(\bar{x}_j - \left(\frac{\sum_{j=1}^{c}\frac{n_j}{s_j^2}\bar{x}_j}{\sum_{j=1}^{c}\left(\frac{n_j}{s_j^2}\right)}\right)\right)^2}{1 + \frac{2(c-2)}{c^2-1}\sum_{j=1}^{c}\left(\frac{1}{n_j-1}\right)\left(1 - \frac{\left(\frac{n_j}{s_j^2}\right)}{\sum_{j=1}^{c}\left(\frac{n_j}{s_j^2}\right)}\right)^2}$$

where c is the number of groups, n_j is the sample size of the j-th group, s_j^2 is the sample variance of the j-th group, and $\bar{x}_j$ is the arithmetic mean of the j-th group. The test follows an $F(1-\alpha, df)$ distribution, where the degree of freedom, df, is:

$$df = \frac{c^2 - 1}{3\sum_{j=1}^{c}\left(\frac{1}{n_j-1}\right)\left(1 - \frac{\left(\frac{n_j}{s_j^2}\right)}{\sum_{j=1}^{c}\left(\frac{n_j}{s_j^2}\right)}\right)^2}$$

The Kruskal-Wallis test is used when the normality assumption is violated (LEVINE et al., 2016; WEISBURD & BRITT, 2014; DAALGARD, 2008; DAWSON & TRAPP, 2003; SIEGEL, 1979). The van der Waerden test is obtained as follows (NIST, 2015; CONOVER, 1999):

$$T_1 = \frac{1}{s^2}\sum_{i=1}^{c} n_i(\bar{A}_i)^2$$

where n_i is the sample size of the group, s^2 is the variance of normal scores and $\bar{A}_i$ is the mean of normal scores. The normal scores, the average of the normal scores and the variance are obtained from the following equations (CONOVER, 1999):

$$A_{ij} = \frac{\phi^{-1}\left(R(x_{ij})\right)}{N+1}$$

$$\bar{A}_i = \frac{1}{n_i}\sum_{1}^{n_i} A_{ij}$$

$$s^2 = \frac{1}{N+1}\sum_{i-1}^{c}\sum_{j=1}^{n_i} A_{ij}^2$$

where N is the total sample size, ϕ^{-1} is the normal percentage point function and $R(x_{ij})$ is the observation rank. This test is compared with a critical value, which follows a chi-square distribution, with $k-1$ degrees of freedom $\chi^2(1-\alpha; k-1)$. The null hypothesis is that the k distributions are equal. If the null hypothesis is rejected, then a post hoc test is applied to determine which pair(s) of means differ. If the difference, in module, of the average normal scores is greater than the critical range, then there is a significant difference between those pairs (CONOVER, 1999):

$$|\bar{A}_i - \bar{A}_j| > t\left(1 - \frac{\alpha}{2}, N-k\right)\sqrt{s^2\frac{N-1-T_1}{N-k}}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

RESULTS AND DISCUSSION

For the case study with simulated data, the ANOVA table was obtained (Table 2). As the value of the calculated F statistic was greater than the critical value, the null hypothesis of equality of population means was rejected. Normality was inspected using the Shapiro-Wilk and Jarque-Bera tests (Table 3). The variables *Trat2* and *Trat3* were rejected in the null hypothesis in the Shapiro-Wilk test, but not in the Jarque-Bera test. Asymmetry is a measure of the lack of symmetry of a given set about the mean. This information is relevant to the point that when the set exposes outliers, the data distribution can exhibit some kind of asymmetry. However, a null skewness does not necessarily imply that the distribution is symmetric, but rather that the values are distributed relatively equally on both sides of the mean. Thus, Figure 1 presents the box plots for the case study with simulated data. The first graph shows that the variables *Trat2* and *Trat4* each have one outlier. Although there are suspicious values, they are not above three standard deviations, as can be seen in the graph on the right of Figure 1.

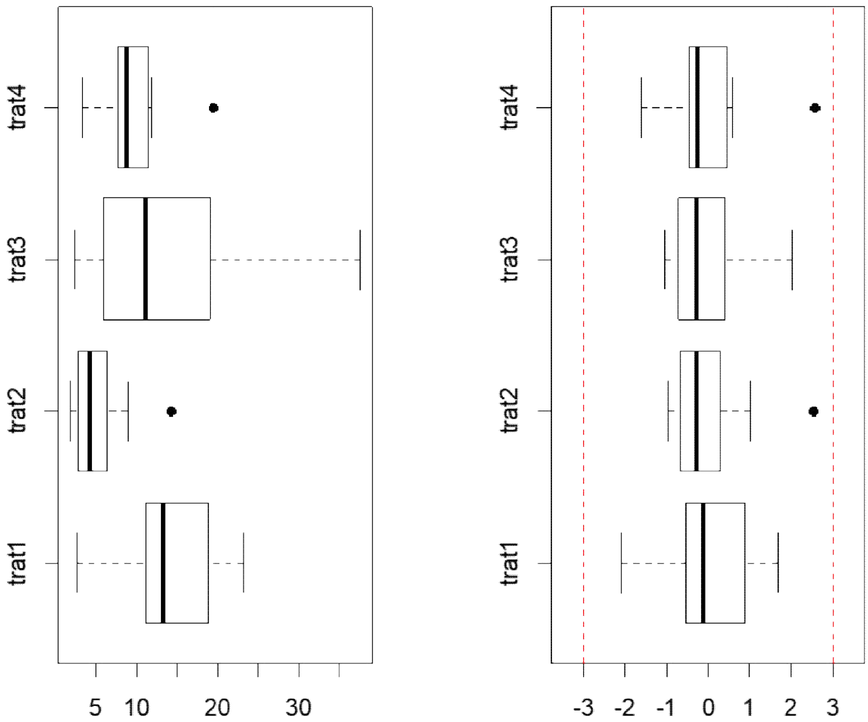

Figure 1. Box-plots for variables

Table 1. Simulated data

Sample	Trat1	Trat2	Trat3	Trat4
1	9,448805	2,329382	12,829105	3,377990
2	18,346060	5,483498	8,220362	7,106937
3	23,182850	8,841417	20,577545	11,758749
4	2,624464	4,827907	10,813706	10,996096
5	19,255240	3,324504	16,974198	19,508620
6	19,716324	1,823120	11,349462	8,317335
7	13,238621	2,100836	23,800015	9,265200
8	13,407086	14,305197	14,161527	11,819506
9	13,300282	5,744728	12,733081	8,647556
10	11,348897	3,400752	14,004753	8,762333
11	13,823267	3,588856	7,115411	8,145451
12	10,699515	6,965917	9,934310	7,502509

Table 2. ANOVA

Source of Variation	df	SS	MS	F-value
Between groups	3	657,42	219,14	4,66
Within groups	44	2070,98	47,07	
Sum	47	2728,39		

Table 3. Test results (p-values)

Group	Shapiro-Wilk	Jarque-Bera
Trat1	0,730	0,929
Trat2	0,027	0,063
Trat3	0,047	0,368
Trat4	0,059	0,090

Table 4. ANOVA of the Brown-Forsythe test

Source of Variation	df	SS	MS	F-value
Between groups	3	235,91	78,64	3,01
Within groups	44	11449,09	26,12	
Sum	47	1385,01		

Table 5. Tukey-Kramer test

Group 1	Group 2	Mean	Standard Error	p-value
Trat1	Trat2	8,805	1,980	0,015
Trat1	Trat3	0,210	1,980	0,999
Trat1	Trat4	4,432	1,980	0,399
Trat2	Trat3	9,015	1,980	0,013
Trat2	Trat4	4,373	1,980	0,410
Trat3	Trat4	4,642	1,980	0,358

Thus, the next step is to verify the homoscedasticity of the data, using the Brown-Forsythe test (Table 4). As the value of the F test is greater than the critical one (3.01 > 2.82), the null hypothesis of homoscedasticity is rejected. In this case, Welch's ANOVA can be used. Welch's ANOVA provided the F test equal to 8.28. The critical F is equal to 3.03. The p-value was less than 0.001, which indicates that at least one mean is different from the others. When applying the Tukey-Kramer test, the pairs of means that were marked as different are *Trat1 vs Trat2* and *Trat2 vs Trat3*. The results are shown in Table 5. The use of the van der Waerden test found a calculated value of T_1 equal to 14.28 and (p-value equal to 0.002). The investigation to see which mean pairs differ indicates *Trat1 vs Trat2* and *Trat2 vs Trat3*. In undergraduate courses in Brazil, the content of experimental statistics is briefly explored, forgetting some good methodologies. Most institutions choose to explain only the one-way ANOVA and, when it does not meet the necessary assumptions, the indication is to use the Kruskal-Wallis test. The Brazilian institutions fail to demonstrate other resources and techniques. There are other methods that can be applied, such as the use of Welch's ANOVA (which is a little more robust) and the van der Waerden test. As a proposal for future works, the suggestion is to use Welch's ANOVA and the van der Waerden test on real data, such as policy public (violence, health, education, environment, and others).

CONCLUSION

In this work, from simulated data, it was found that Welch's ANOVA and the van der Waerden test can be used with good results. Therefore, their investigation could be more used and deepened by the professors of the Brazilian institutions.

REFERENCES

BLACK, Ken. Business statistics: for contemporary decision making. John Wiley & Sons, 2019.
BRANDT, S. Data Analysis: Statistical and Computational Methods for Scientists and Engineers, 4th ed. Springer, 2014.
CONOVER, W. J. Practical nonparametric statistics, 3rd ed. New York: John Wiley & Sons, 1999.
DAALGARD, P. Introductory Statistics with R, 2nd ed. New York: Springer. 2008.
DAWSON, B.; TRAPP, R. GBasic and Clinical Biostatistics, 3ªed. McGraw-Hill, 2003.
FONTELLES, M. J. Biostatistics applied to operations research, v. 2. São Paulo: Livraria da Física, 2012.
HEIBERGER, R. M.; HOLLAND, B. Statistical Analysis and Data Display: An Intermediate Course with Examples in R, 2nd ed. Springer, 2015.
LEVINE, et al. Statistics: Theory and applications using MS-Excel in Portuguese, 7ª ed. LTC, 2016.
MENDENHALL, W. Probability and statistics, v. 2. Campus, 1985.
NAGADEVARA, V. Statistical Methods: Basic Inferences. In: Essentials of Business Analytics: An Introduction to the Methodology and its Applications. Cham: Springer, 2019.
NIST. Anderson-Darling and Shapiro-Wilk tests. Engineering Statistics Handbook. 2015. https://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm.
NIST. Van der Waerden. NIST Dataplot. 2015. https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/vanderwa.htm.
HILTON, G. R.; STREINER, D. L. Biostatistics: the bare essentials, 3rd ed. Hamilton: B.C. Decker, 2008.
R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. https://www.R-project.org.
SIEGEL, S. Nonparametric statistics for the behavioral sciences, 2ª ed. McGraw-Hill, 1979.
SPIEGEL, M. R. Schaum Collection: Probability and Statistics.Rio de Janeiro: McGraw-Hill do Brasil, 1979.
SPRENT, I. P.; SMEETON, N. C. Applied nonparametric statistical methods. Boca Raton: CRC Press, 2001.
STOCKEMER, D. Quantitative Methods for the Social Sciences: A Practical Introduction with Examples in SPSS and Stata. Cham: Springer, 2019.
TRIOLA, M. F. Introduction to Statistics, 10ª ed. Rio de Janeiro: LTC, 2008
VIEIRA, S. Biostatistics: advanced topics, 3ª ed. Rio de Janeiro: Elsevier, 2010.
WALPOLE, R. E. et al. Probability and Statistics for Engineering and Science, 8ª ed. São Paulo: Pearson Prentice Hall, 2009.
WEISBURD, D.; BRITT, C. Statististics in Criminal Justice, 4th ed. New York: Springer, 2014.
WONNACOTT, T. H.; WONNACOTT, R. J. Introduction to Statistics.Rio de Janeiro: LTC, 1980.
ZAIONTZ, C. Welch's ANOVA Test.Real Statistics using Excel. 2020. https://www.real-statistics.com/one-way-analysis-of-variance-anova/welchs-procedure.
ZAIONTZ, C. REGWQ Post-hoc test. 2020. http://www.real-statistics.com/one-way-analysis-of-variance-anova/unplanned-comparisons/regwq-post-hoc-test/#:~:text=The%20Ryan%2C%20Einot%2C%20Gabriel%2C%20Welsh%20Studentized%20Range%20Q,order%20x%CC%841%20%E2%89%A5%20x%CC%842%20%E2%89%A5%20%E2%80%A6%20%E2%89%A5%20x%CC%84k.
ZAR, J. H. Biostatistical analysis, 5th ed. Essex: Pearson Education, 2014.

****