

RESEARCH ARTICLE

A web-based software developed for permutation tests and an application in medicine

Zeynep Tunç¹, Şeyma Yaşar¹, Emek Gldođan¹, Cemil olak¹

¹Department of Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya, Turkey

Received: 16 March 2020, Accepted: 05 April 2020, Published online: 31 August 2020

© Ordu University Institute of Health Sciences, Turkey, 2020

Abstract

Objective: In this study, it is aimed to develop a new user-friendly web-based software in order to easily carry out the use of permutation tests that can overcome the difficulties of use due to the restrictions in the usage phases of parametric and nonparametric tests and can be used as an alternative to these tests.

Methods: Shiny, an R package, was used to develop the "permutation tests" software. In the developed software, by selecting "the Specify Sample Number" tab, the number of samples presented as "Single", "Two" and "More than two" options is selected and analyzes are made by selecting the appropriate data set from the file upload menu.

Results: The data set called "dietstudy" was used to examine the work of the developed web-based software and to evaluate its outputs. "Two Independent Sample Permutation Tests" were selected and analyzed to see whether there was a difference between the variables in terms of gender. According to the results, no statistically significant difference was found for the triglyceride levels Triglyceride, 1st interim triglyceride, 2nd interim triglyceride, 3rd interim triglyceride ve Final triglyceride in terms of gender, but a statistically significant difference was obtained in terms of Weight, 1st interim weight, 2nd interim weight, 3rd interim weight ve Final weight variables.

Conclusion: The "permutation tests" software developed is a new user-friendly web-based software that can be used to easily perform permutation tests that can be used as an alternative to the preferred parametric and non-parametric tests.

Key words: Parametric tests, nonparametric tests, permutation tests, Web-based software

Suggested Citation: Tunc. Z. Yasar. Ş. Guldogan.E Colak. C. A web-based software developed for permutation tests and an application in medicine. Middle Black Sea Journal of Health Science, 2020; 6(2):207-211.

Address for correspondence/reprints:

Zeynep Tun

Telephone number: +90 422 341 0660/1281

ORCID-ID 0000-0001-7956-9772

E-mail: zeynep.tunc@inonu.edu.tr

DOI: 10.19127/mbsjohs.704457

Note: This study was presented as oral presentation in ISMSIT 11-13 October 2019 - Ankara (3rd International Symposium on Multidisciplinary Studies and Innovative Technologies).

Introduction

Statistical hypothesis tests are used in order to reach the results suitable for the purpose determined by making statistical inferences with the data obtained as a result of the studies. Statistical tests used to analyze the data can be classified as parametric and nonparametric tests according to the structure of the data, the type of scale, the form of the distribution and the hypothesis to be tested. Parametric tests are methods based on certain hypothetical population distribution and some parameters such as mean, a standard deviation that determine this distribution. Nonparametric tests are generally based on more basic estimations such as expected-observed value differences, sequence and sequence differences, independent of population distribution (Buskirk, Willoughby and Tomazic 2013). Statistical power and effect size of parametric tests are higher than nonparametric tests (Kartal 2010). For this reason, researchers tend to choose and apply parametric tests rather than non-parametric tests in the selection of tests (Shapiro and Wilk 1965). Parametric tests contain some assumptions since they are developed according to a known distribution and some parameters. Some of these assumptions show that the universes from which the samples are taken have a normal distribution and the variance of the universe is homogeneous. (Alpar 2013). Neglecting assumptions for these tests can cause test results to be inaccurate and misleading. In such cases, it is ensured that the data set is made to meet the assumptions by making necessary arrangements (for example, using data conversion methods if it is not normally distributed). Use of non-parametric tests is preferred if assumptions are not met in spite of the procedures (Field 2013). In non-parametric tests, operations can generally be performed by considering the ordinal numbers. Therefore, the use of permutation tests which can be used in place of both parametric and nonparametric tests provides different advantages. Because, as in parametric tests, it does not require any assumption about sample distribution. In addition, permutation tests use the actual values of the data, not the ordinal numbers of the data, as in nonparametric tests. When compared in terms of test results, it is reported that permutation tests are close to parametric tests, sometimes they give better results, and often they are better than non-parametric tests (Good 2006).

Although the use of the permutation method was first proposed at the beginning of the twentieth century, it has become more preferred in recent years with the elimination of the difficulty in calculations with the development and use of computer

technology (Good 2006). Permutation tests have been preferred nowadays due to the limitations in the use of parametric and nonparametric tests (Ludbrook and Dudley 1998). For this reason, it is aimed to develop a new easy-to-use web-based software to carry out the use of permutation tests that can be used to avoid restrictions.

Methods

In order to examine the working principle of the web-based software and evaluate its outputs, the data set called dietstudy was obtained from the link https://www.ibm.com/support/knowledgecenter/ko/S SLVMB_23.0.0/spss/tutorials/data_files.html. This data set contains the results of a study of the "Stillman diet". In the data set, age, sex variables and weights of individuals before and after the diet were given in weight and triglyceride levels were given in mg / 100 ml.

Permutation test

Although the permutation test method was first introduced and proposed in the early 20th century, it has been used more and more recently with the elimination of the difficulty in calculations due to the development of computer technology (Good 2006). Permutation tests are increasingly common tests to perform certain types of statistical analysis. As with other known parametric and non-parametric tests, permutation tests do not rely on assumptions about the distribution of data and are therefore considered non-parametric tests (Mangiafico 2016).

Permutation tests have certain advantages over parametric and non-parametric tests for ease of use (Ludbrook and Dudley 1998). Permutation tests do not require some assumptions as in parametric tests. For example, the assumption about the distribution of samples which is one of the parametric test assumptions does not exist in the permutation tests. Similarly, non-parametric tests are generally performed with the ordinal numbers of data, and in permutation tests operations are performed using the actual values, not the ordinal numbers of the data. Permutation tests give results that are close to parametric test results, but often give better results when it comes to non-parametric test results. For these reasons, the permutation test method can be used instead of all parametric and non-parametric test methods (Good 2006).

Developed web-based software

To create this web-based application, the Shiny library, which allows the design of interactive web-based applications, is used (Chang et al. 2017). The

main and sub-menus created to use the software are described below.

File upload

In the analysis to be made with this application, the first step is loading the file containing the data set. Data analysis is done by loading MS Excel (.xls / .xlsx) and SPSS (.sav) file types, which have different extensions and are widely used. In addition, this menu contains the Specify the type and role of the variables tab that will allow you to specify the type and role of the variables in the loaded file. In addition, this menu contains the "Specify the Type and Role of Variables" tab, which will allow you to specify the type and role of the variables in the uploaded file. This process is necessary for the analysis and the roles of the variables should be determined.

Permutation tests

In this menu of web-based applications, the number of samples should be selected in order to decide the test to be used. When the permutation tests menu is opened, the number of samples presented with the option "One", "Two" and "More than Two" is selected from the "Number of Sample" tab. In this tab, "One Sample Permutation Tests" for "Single" number of sample, "Two Independent Sample Permutation Tests" and "Two Dependent Sample Permutation Tests" for "Two" number of sample and

"More Than Two independent Sample Permutation Tests" and "More than Two Dependent Sample Permutation Tests" for "More than Two" number of sample tabs are opened. The number of samples appropriate to the loaded data set is determined and analyzes are performed with appropriate test selections. Figure 1 shows the "Permutation Tests" menu.



Figure 1. "Permutation Tests" menü

Results

In order to examine the working principle of the developed web-based software and evaluate its outputs, the data set called dietstudy was used. A descriptive statistics table related to the data set is given in table 1.

First, the data set is loaded into the software. Images of the data set uploaded to the software are shown in Figure 2.

Table 1. Descriptive statistics table of variables

Variables	Variable Type	Mean±Standart deviation
Triglyceride	Numerical	138,44±29,040
1st interim triglyceride	Numerical	124,56±25,126
2nd interim triglyceride	Numerical	124,38±21,854
3rd interim triglyceride	Numerical	118,81±33,255
Final triglyceride	Numerical	124,38±29,412
Weight	Numerical	198,38±33,472
1st interim weight	Numerical	196,13±33,669
2nd interim weight	Numerical	194,13±33,498
3rd interim weight	Numerical	192,13±33,915
Final weight	Numerical	190,31±33,508
Age in years	Numerical	54,69±6,916

Then, with the "Permutation Tests" tab, the appropriate test selection is made according to the purpose of the study and the data set. "Two Independent Sample Permutation Tests" were selected in order to see whether there was a difference between the variables in terms of gender and the analyzes were performed. Figure 3 shows the images for a variable belonging to the results of "Two Independent Samples Permutation Tests".

According to these results, no statistically significant difference was found for Triglyceride, 1st interim triglyceride, 2nd interim triglyceride, 3rd

interim triglyceride and Final triglyceride variables. A significant difference was obtained in terms of weight, 1st interim weight, 2nd interim weight, 3rd interim weight, and Final weight variables.

A web-based software developed for permutation tests

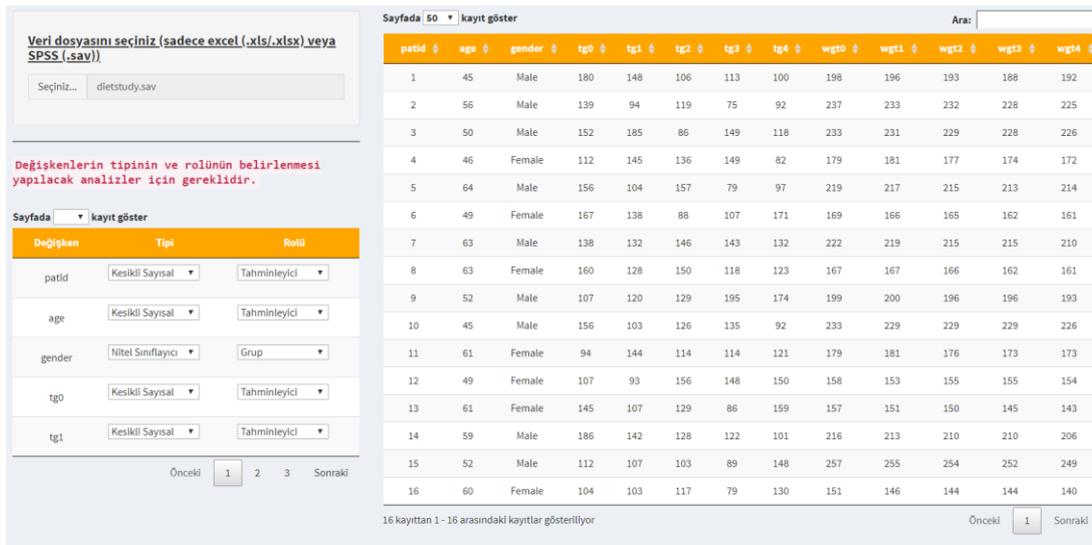


Figure 2. The images of the uploaded file.

Değişken adı: tg3

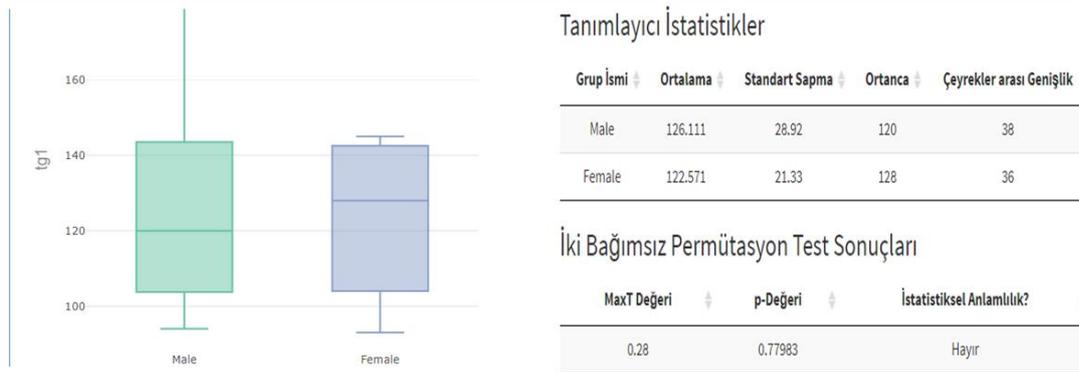


Figure 3. “Two Independent Sample Permutation Tests” result

Discussion

In this study, it is aimed to develop a user-friendly web-based software to easily carry out the use of permutation tests in order to overcome the usage difficulties due to the restrictions in the use of parametric and non-parametric tests which are generally preferred tests. Parametric tests are methods based on a certain hypothetical population distribution and some parameters such as mean, standard deviation that determine this distribution. Nonparametric tests are generally based on more basic estimations such as expected-observed value differences, sequence and sequence differences, provided that they are independent of population distribution (Buskirk et al. 2013). Since parametric tests are developed based on a certain distribution and parameters, they contain some assumptions that are defined based on these limitations (Alpar 2013). The

use of parametric tests depends on the fulfillment of these assumptions. Nonparametric tests perform operations by using sequence numbers of data (Good 2006). Due to such similar restrictions and conditions, it is advantageous to use permutation tests that can be used in place of the proposed parametric and nonparametric tests at the beginning of the 20th century. (Ludbrook and Dudley 1998). Because it does not require any assumption about sample distribution which is one of the parametric test assumptions. In addition, as in non-parametric tests, the data is processed using real values, not sequence numbers (Good 2006). Therefore, a user-friendly web software has been developed in which researchers can easily access permutation tests and interpret their results easily.

IBM SPSS Statistics (Corp 2017), Minitab (Minitab 2000), MedCalc (Schoonjans et al. 1995),

Stata (StataCorp 2007), which are known package programs, do not have a module to allow the application of permutation tests. Using web-based software developed in this study, comparisons of permutation tests can be done easily by loading the data set and interpretation can be obtained for statistical significance.

The developed software in accordance with the number and structure of the sample includes Single Sample Permutation Tests, Two Independent Sample Permutation Tests, Two Dependent Sample Permutation Tests, More Than Two Independent Sample Permutation Tests, and More than Two Dependent Sample Permutation Tests. The software provides test results, descriptive statistics of variables and graphs of variables.

In addition, the data set called diet study was used in order to examine the working principle of the web-based software developed in this study and to evaluate its outputs. According to the results obtained by using “Two Independent Sample Permutation Tests in order to see whether there is a difference between the variables in terms of gender, no statistically significant difference was found for Triglyceride, 1st interim triglyceride, 2nd interim triglyceride, 3rd interim triglyceride and Final triglyceride variables in terms of gender. However, a significant difference was obtained in terms of Weight, 1st interim weight, 2nd interim weight, 3rd interim weight and Final weight

Conclusion

Permutation tests have been preferred nowadays due to the limitations in the use of parametric and nonparametric tests (Ludbrook and Dudley 1998). For this reason, a new user-friendly web-based software has been presented to users in order to realize the use of permutation tests that are used as an alternative to parametric and nonparametric tests.

Ethics Committee Approval: Ethics committee approval is not required in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept- Z.T, S.Y; Design- E.G.C.C; Materials- Z.T, S.Y, E.G; Data Collection and Processing- Z.T.; Literature Review- S.Y, E.G, C.C; Writing- Z.T, S.Y.; Critical Review- Z.T.C

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study hasn't received no financial support.

References

- Alpar R. Applied multivariate statistical methods. Detay Publishing, 2013.
- Buskirk T D, Willoughby L M, Tomazic T J. Nonparametric statistical techniques. The Oxford handbook of quantitative methods. Statistical analysis. 2013;2:106-141.
- Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. R package version 1. 2017.
- Corp, I. IBM SPSS statistics for windows, version 25.0. Armonk, NY: IBM Corp. 2017.
- Field A. Discovering statistics using IBM SPSS statistics. Sage. 2013.
- Good P I. Resampling methods. Springer. 2006.
- Hothorn T, Hornik K, Van De Wiel M A, Zeileis A. Implementing a class of permutation tests: the coin package. 2008.
- Kartal M. Hypothesis Tests in Scientific Research. Nobel publishing. Ankara. 2010.
- Ludbrook J, Dudley H. Why permutation tests are superior to t and F tests in biomedical research. The American Statistician. 1998;52:127-132.
- Mangiafico S S. Summary and analysis of extension program evaluation in R. Rutgers Cooperative Extension: New Brunswick. NJ, USA. 2016.
- Minitab I MINITAB statistical software. Minitab Release. 2000.
- Schoonjans F, Zalata A, Depuydt C, Comhaire F. MedCalc: a new computer program for medical statistics. Computer methods and programs in biomedicine. 1995;48:257-262.
- Shapiro S S, Wilk, M B. An analysis of variance test for normality (complete samples). Biometrika. 1965;52:591-611.
- StataCorp L. Stata data analysis and statistical Software. Special Edition Release. 2007;10:733.