

A Comprehensive Guide for Selecting Appropriate Statistical Tests: Understanding When to Use Parametric and Nonparametric Tests

Saed Jama Abdi 

Norwegian Refugee Council (NRC), Mogadishu, Somalia

Email: saedjamatr@gmail.com

How to cite this paper: Abdi, S.J. (2023) A Comprehensive Guide for Selecting Appropriate Statistical Tests: Understanding When to Use Parametric and Nonparametric Tests. *Open Journal of Statistics*, 13, 464-474.

<https://doi.org/10.4236/ojs.2023.134023>

Received: June 15, 2023

Accepted: August 5, 2023

Published: August 8, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Choosing appropriate statistical tests is crucial but deciding which tests to use can be challenging. Different tests suit different types of data and research questions, so it is important to choose the right one. Knowing how to select an appropriate test can lead to more accurate results. Invalid results and misleading conclusions may be drawn from a study if an incorrect statistical test is used. Therefore, to avoid these it is essential to understand the nature of the data, the research question, and the assumptions of the tests before selecting one. This is because there are a wide variety of tests available. This paper provides a step-by-step approach to selecting the right statistical test for any study, with an explanation of when it is appropriate to use it and relevant examples of each statistical test. Furthermore, this guide provides a comprehensive overview of the assumptions of each test and what to do if these assumptions are violated.

Keywords

Statistical Tests, Levels of Measurement, Parametric, Nonparametric, Normal Distribution

1. Introduction

Statistics is the science of how to collect, organize, analyze, and interpret numerical information in order to draw relevant conclusions. For each situation, there are statistical tests appropriate for the analysis and interpretation of the data. Meaningful conclusions can only be drawn based on the data collected us-

ing right statistical test. However, choosing appropriate statistical test is very important but also very difficult. Selecting inappropriate statistical test create some severe problems during the interpretation and also lead to erroneous conclusions of the study [1]. Using incorrect statistical tests can result in inaccurate conclusion and it can have serious consequences for the validity and reliability of the research findings. The accessibility of distinct types of statistical packages such as R, Stata, Minitab, SPSS etc, makes performing statistical tests easily, but selection of the proper statistical test is still a difficult task because software does not have the ability to choose correct statistical test for a given situation. To have the knowledge of choosing correct statistical test is essential for making informed decisions, generating reliable findings and identifying patterns and trends. The researcher selects appropriate statistical test based the nature of the collected data and objective of analysis [2]. To select right statistical test one should need to know basic knowledge of statistics such conditions and assumptions of statistical tests that will help to select correct statistical analysis during the data analysis. The aim of this paper is to present a systematic method and guide junior researchers to easily choose appropriate statistical methods. In order to choose appropriate statistical test one should ask the following questions.

What is the objective of the analysis?

What are the levels of measurement of the data?

Is the data parametric or nonparametric?

1.1. Objectives of the Analysis

To select appropriate statistical test, it is crucial to determine the objective of the analysis. The researcher must make sure that the statistical test is relevant for the way that the research was designed and the kind of data collected [3]. We should ask, what are we looking for? Are your looking for the relationship between two variables, for example, the relationship between sales and advertisement, then in that case you require to use correlation analysis.

If the objective of the study is to compare the means of three or more groups, you need to use analysis of variance (ANOVA), if you are interested in comparing two groups you require to use two-sample test whether it is paired test or independent test. Suppose we want to know if the mean monthly income of employees from a large company is the same or different from \$500, then in this case the most appropriate statistical test is to use hypothesis test for single population mean. If you are looking for to test whether a sample proportion is statistically different from population proportion, then chi-square test, Fisher's exact test or McNemar's test is appropriate.

1.2. The Scales of Measurement

After determining objective of the study, the next step is to know what scales of measurement for the data are. To know scale of measurement of data involves a researcher must understand the types of the variable that involves the data anal-

ysis. In general, variable can be either quantitative or qualitative (As shown **Figure 1**). Quantitative variable is a variable that can be measured in numbers such as age, income, number of customers visited in a bank etc, this types variable can either interval or ratio scale of measurement. Qualitative variable is a type of variable that cannot be measured in numbers like sex, education level, and marital status can be classifies as either nominal or ordinal scale of measurement.

1.2.1. Nominal Scale of Measurement

If the researcher wants to classify variables based on their names or labels and do not present any order or rank, like to classify respondents based on their sex such as male and female or classify according to their preference for a particular product like prefer and do not prefer then level of measurement used is nominal scale. The important point about nominal level of measurement is that there are no ranks among the responses. For example when classifying respondents according to their gender is meaningless to say that male is placed a head of female, responses are classified only by name. If the data consists of two groups such as male and female or prefer and do not prefer are known as binomial data. However, if it consists of more than two classes it is called multinomial data. These type of data usually summarized in the form of contingency table.

1.2.2. Ordinal Scale of Measurement

In ordinal level of measurement, the responses can be ranked, however the difference between the ranks are meaningless. Unlike nominal, ordinal level of measurement can be ranked on the base of some relationship among them. For example, a marketing researcher wishing to know satisfaction of customers might ask them their feeling towards the services they received as either very dissatisfied, dissatisfied, satisfied, or very satisfied. The responses in this question is ranked in an order, ranges from the least satisfied to most satisfied. Ordinal level of measurement, we can simply tell that one response is equal to, less than or greater than another. All Likert scale questions are good example of ordinal level of measurement.

1.2.3. Interval Scale of Measurement

Data with interval scale of measurement is quantitative data that has a meaningful

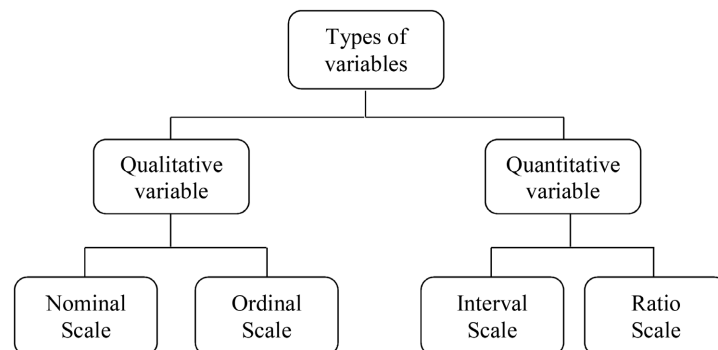


Figure 1. Flow chart for selecting scales of measurement.

order between the ranks like age in years, weight (kg) or distance travelled by customers to buy a specific product. An interval scale dataset has the property that equal intervals between measurements reflect equal changes in quantity [1]. However, ordinal scale data has no true zero. For example in temperature, we cannot say 40°C is twice as hot as 20°C or 0°C does not mean there is no temperature. Another good example of interval data is the IQ test.

1.2.4. Ratio Scale of Measurement

The most informative scale of measurement is the ratio level of measurement. Ratio scale data have the attributes of the interval scale data, however, additional property that these types of data has is that it has zero meaningful. Money is a good example of ratio scale, for example, one employee has a monthly salary of \$4000 and another has \$2000. Since money has a true unit, we can say that an employee with \$4000 is twice as much as an employee with \$2000 salary. Another good example of these types of data is weight (Kg). For example, if someone can lift 100 kg and another can lift 50 kg. We can say that the first person can lift twice as much as second person.

1.3. Parametric and Nonparametric Methods

1.3.1. Parametric Methods

Parametric methods are power, however they require to test different assumptions, especially for interval or ratio scale of measurement data, it is important to examine the distribution of the data to check if it is normally distributed—that is symmetric around the mean. The Histogram in **Figure 2** is an example of normally distributed data, while the one in **Figure 3** is an illustration of skewed

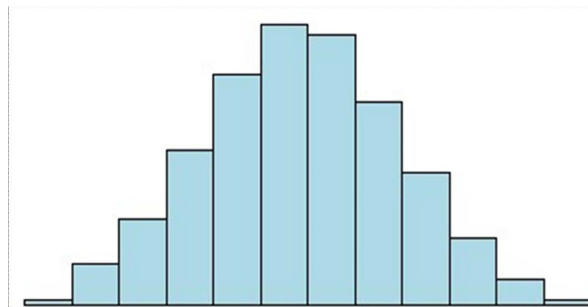


Figure 2. Normal distribution.

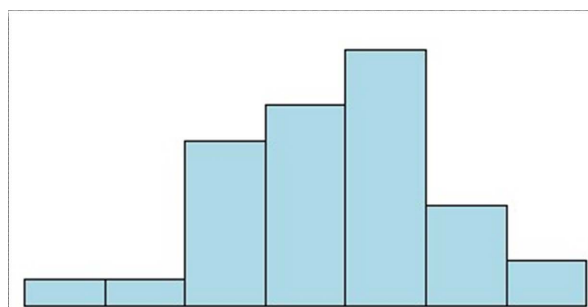


Figure 3. Not normal distribution.

data. Instead of using Histogram, there are various different methods of test of normality like, Normal Q-Q plot, Shapiro-Whilks test, Kolmogorov-Smirnov test, and Box plot.

1.3.2. Nonparametric Methods

Non-parametric tests are distribution free methods, which do not require to meet the required assumptions such normal distribution. Unlike parametric tests, these distribution free tests can be used both qualitative and quantitative data. Parametric statistical methods require assumptions such as normality of the population from which sample(s) are drawn. However, if these assumptions are violated a transformation method can be applied [4].

2. Testing Population Parameter Theory

2.1. One Sample T-Test

The One sample t-test is used to compare one sample to a hypothesized value or to test a certain given theory about a population parameter for a ratio or interval scale of measurement variable, where the distribution is approximately normally distributed or where the sample size is sufficient ($n \geq 30$).

When using a one-sample t-test, the researcher hypothesizes a specific value for the mean of a population and then selects a sample of data to compare the hypothesized value with. For instance, you want to test the claim that the mean salary of computer programmers in a certain country is at least \$30,000 annually.

2.2. The Sign Test

The sign test for single samples is used to test the value of a median for a specific sample for an ordinal level of measurement variable and/or when the condition of normality cannot be met. When using the sign test, the researcher hypothesizes the specific value for the median of a population; then he/she selects a sample of data and compares each value with the conjectured median.

If the data value is above the hypothesized median, it is assigned a plus sign, and if it is below, it is assigned a minus sign, while those that are the same as the conjectured median are assigned a 0.

The null hypothesis is true, if the number of plus signs are approximately equal to the number of minus signs. If the null hypothesis is not true, there will be a disproportionate number of plus or minus signs.

3. Comparing Two Groups

3.1. The Unpaired Samples T-Test

The unpaired samples t test, also known as the independent samples t test, is used for comparing the means of two independent sets of data. It is a parametric test, meaning that it assumes that the data follow a normal distribution. The test also assumes that the two samples are independent of each other.

For example, suppose you want to measure the salary of 200 senior managers:

100 female and 100 male and you want to compare if the mean salary of female managers is significantly different from that of the male when the salary is approximately normally distributed.

3.2. Mann-Whitney U test

The Mann-Whitney U test is a nonparametric version of the unpaired samples t-test, which applies to comparing two independent groups of an ordinal or interval data when the assumptions for the unpaired samples t-test (normality) do not met—that is, the Mann-Whitney U test does not require the assumption of normality. Unlike the unpaired sample t-test, the Mann-Whitney is used to test the hypothesis that the medians of two sets of data are equal.

For example, to compare the readability of two groups of students, a control group and a treatment group, when the readability score is not normally distributed.

3.3. Chi-Square (χ^2) Test for Contingence Table

The chi-square test is used to examine if there is an association between two categorical (nominal level of measurement) variables presented in a cross-tabulation table with one variable as a row and as a column for the other variable. However, the assumption of this test is the sample must should large ($n > 20$) and the two variables are unrelated. If the sample size is small, that is less than 20 in a 2 by 2 cross-tabulation table, but the samples are independent, then Fisher's exact test is used [5].

For example, you want to investigate a group of customers to know what type of computer brand they prefer. The attribute is the sex of the customer (male or female), and the dependent variable is the preferred brand (HP or Samsung) (Figure 4).

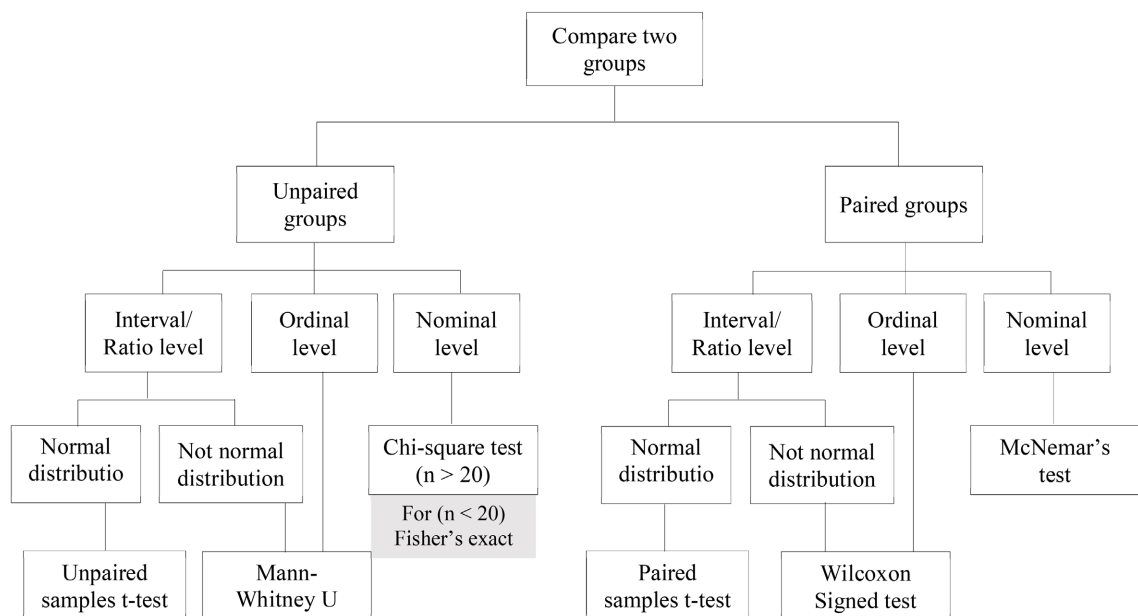


Figure 4. Choosing best statistical tests for comparing two groups.

3.4. Paired Samples T-Test

The paired samples t-test, also known as the dependent samples t-test, compares the means of two related groups to test whether there is a significant difference between them. It is applicable when the distribution is normally (approximately) distributed and the data is either interval or ratio scale of measurement.

For example, suppose human resources department of a company wants to evaluate the effectiveness of a training program. One way of doing this is to measure the performance of the employees before and after the training.

3.5. Wilcoxon Signed Test

The Wilcoxon signed rank is a nonparametric test equivalent to the paired samples t-test, which is used to compare two paired samples when data is either interval scale but assumptions for the paired sample *t*-test are not met or when data is *ordinal* scale of measurement.

Unlike the mean difference in the paired sample t-test, the hypothesis being tested is whether the median difference is zero.

For example, to test the performance of sales persons before and after training (when data is either not normal distribution or the sample size is small). In this example, the dependent variable is sales performance and the two-paired groups sales performance before and after training.

3.6. McNemar's Test

The McNemar's test is a nonparametric test that is used to compare the differences between proportions in two paired groups (before and after) when data is at the nominal level of measurement with a binary dependent variable (yes and no) and one independent categorical variable. It is used when the sample size is small and the data is not normally distributed. In this type of test, each participant is measured twice (pre-test and post-test) that represents a repeated measure.

For example, to examine the outcome (purchased or not purchase) of a paired experiment using two different advertised streams (Television or Radio). For more detailed information of McNemar's test [6].

4. Comparing Three or More Groups

4.1. One-Way ANOVA

One-way analysis of variance is a parametric statistical test used to test whether there is a significant difference between the means of three or more groups. It is applicable when you have one categorical independent variable and one quantitative dependent variable. The independent variable should have at least three different groups.

Like any other parametric test, one-way ANOVA assumptions include that samples should be independent, the population from which samples are drawn should be approximately normally distributed, and there should be homogeneity of variance.

Example: You are a bank manager, and you want to test whether or not the 3

tellers in your branch serve, on average, the same number of customers per hour. One way of doing this is to observe and record number of customers served by each of these 3 tellers in 8 hours. Then, run a one-way ANOVA to test whether or not the mean number of customers served per hour by each of the 3 tellers is the same.

4.2. Kruskal-Wallis One-Way ANOVA

The Kruskal-Wallis one-way ANOVA is a non-parametric statistical test used to compare whether or not there is a statistically significant difference between the medians of three or more groups. Unlike one-way ANOVA, the Kruskal-Wallis test is used for ordinal or interval scale variables that are not normally distributed.

For instances, to compare the effectiveness of four different brands of pain relievers (naproxen, ibuprofen, Aspirin, and paracetamol), when data is not normally distributed. Here, the null hypothesis is that there is no difference in pain relief effectiveness between the four brands. The alternative hypothesis is that there is a difference.

4.3. Chi-Square Test of Independence

The Chi-square independence test is a nonparametric statistical test corresponding to one-way ANOVA, which is used to check whether or not two nominal (categorical) variables are related. It is used to analyze nominal data, such as the relationship between gender and iPhone color preference.

Example, to test if there is an association between sex (male or female) and the preferred iPhone color (White, Silver and Gold”). The null hypothesis is that there is no association between sex and iPhone color. The alternative hypothesis is that there is a relationship between these two variables (**Figure 5**).

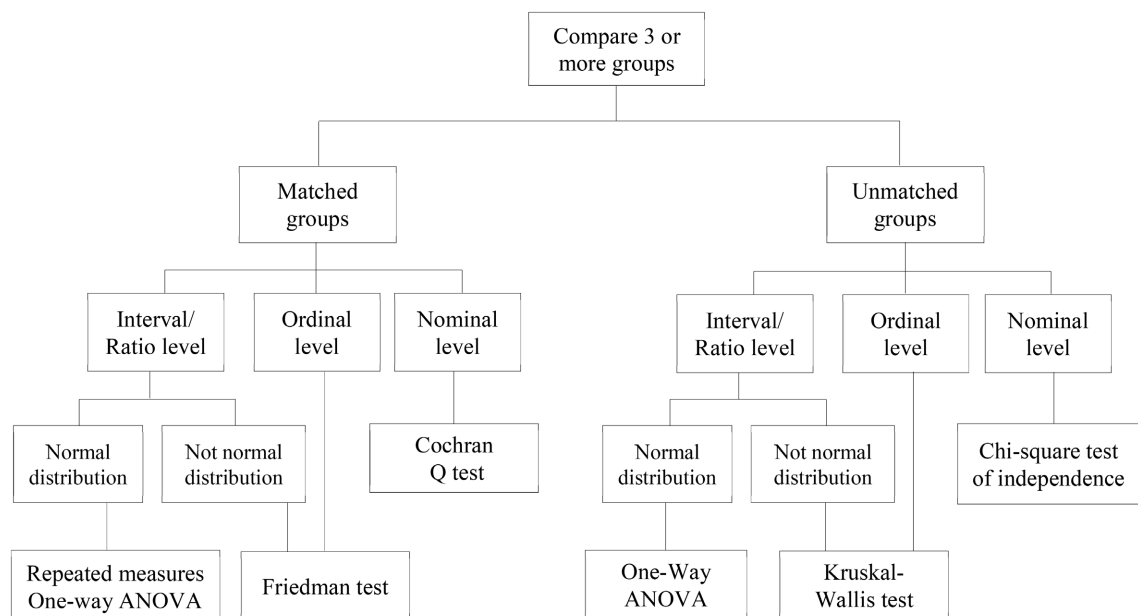


Figure 5. Choosing best statistical tests for comparing three or more groups.

4.4. Repeated Measures One-Way ANOVA

The repeated measures one-way ANOVA is used to test whether there is a statistically significant difference between the means of three or more matched groups when the dependent variable is an interval or ratio (continuous) with an approximately normal distribution and the independent variable is a nominal or ordinal (categorical).

The participants are either the same individuals tested on three or more occasions on the same dependent variable, or the same individuals tested under three or more different conditions on the same dependent variable. This allows researchers to measure changes in the same individuals over time and under different conditions. It also allows them to compare the effects of conditions on the same individuals.

Example: You wish to test the flavor of three different cakes (chocolate, coconut, and carrot) and rate each for its taste. The same people are measured more than once on the same dependent variable (flavor); this is why it is called repeated measures.

4.5. Friedman Test

The Friedman test is a non-parametric statistical test alternative to one-way ANOVA with repeated measures. It is used to examine whether there is a difference between matched (paired) groups when the dependent variable is ordinal, but the assumptions of repeated measures of one-way ANOVA are violated, that is, when data is not normally distributed. The Friedman test is a robust measure that is not affected by violations of the assumption of normality.

For instances, you want to examine whether an advertisement has an effect on customer's decision to buy a product. Here, the dependent variable is customer's decision to buy a product and the independent variable is type of advertisement used (TV, Radio or Magazine). A Friedman test is used to check if there were differences in customer's decision of the product based of the type of advertisement used.

4.6. Cochran's Q Test

The Cochran's Q test is an omnibus nonparametric version of repeated measures one-way ANOVA or an extension of McNemar's test. This test is used when only nominal data is available, and the data analyst wants to determine if there are differences in a dichotomous nominal dependent variable between three or more matched groups. Cochran's Q test has the advantage of not requiring normality and homogeneity of variance assumptions to be met, unlike the repeated measures one-way ANOVA. It is also more powerful than McNemar's test for more than two groups.

Example: To determine whether the percentage of customers who had no rash (as opposed to a rash) increased after three weeks of using a skin cream. At this case, the dependent variable is the effectiveness of the skin cream, which has two

attributes: “rash” and “no rash” measured over three consecutive weeks (“week 1, week 2, and week 3”).

5. Correlation between Variables

5.1. Pearson’s Correlation

The Pearson’s correlation test is used to determine whether there is a linear relationship between continuous (ratio or interval) variables with a normal distribution. When testing the relationship between only two variables, it is a case of simple correlation.

Example: To examine the relationship between monthly household income and expenditure.

5.2. Spearman’s Correlation

The Spearman rank correlation is a nonparametric analog of Pearson’s correlation, which is used to find out the linear correlation between the ranks of two variables. It is used when the assumptions of Pearson’s correlation are not met, that is, when the two variables are not normally distributed.

For example, to examine the relationship between households’ monthly income and expenditure, assuming that both variables (households’ monthly income and expenditure) are not normally distributed.

6. Considerations

In inferential statistics, we typically use either a P-value or confidence intervals to determine whether our statistical results are significant or not. The P-value is a measure of the probability that an observed result is due to chance, while confidence intervals are used to estimate the population parameter based on a sample. Both techniques are used to assess the statistical significance of a result. Although a P-value less than 0.05 is widely used as a cut-off for determining statistical significance, other factors such as the size of the sample have a significant impact on the P-values.

P-values are the probabilities of observing an effect size larger than that detected in a study purely by chance alone. However, it does not indicate the size or direction of the difference. Studies using a large sample size are more likely to achieve statistical significance than their counterparts using smaller sample sizes.

When results are not statistically significant, confidence interval estimates provide more information. The confidence intervals are ranges of possible values for the target population calculated using different statistical methods. 95% is the most commonly used confidence interval, which indicates that if you select 100 samples and calculate a 95% confidence interval for each, then 95 out of 100 intervals will contain true population mean. Unfortunately, like P-values, confidence intervals tend to be reduced with large sample sizes, while they tend to be increased with small sample sizes.

Confidence intervals are generally more informative than P-values because

they provide information on the certainty of the population estimate of interest. They are also easier to interpret than P-values, as they provide a range of possible values for the population parameter of interest.

Finally, statistical results should be interpreted in the context of the research question and study design. When drawing conclusions, researchers should also avoid over-interpreting results and consider other sources of evidence.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Trajkovski, V. (2016) How to Select Appropriate Statistical Test in Scientific Articles. *Journal of Special Education and Rehabilitation*, **17**, 5-28.
<https://doi.org/10.19057/jsr.2016.0>
- [2] Upadhyay, H.P. (2017) How to Choose the Statistical Technique in the Data Analysis. *International Journal of Research Studies in Biosciences*, **5**, 33-37.
<https://doi.org/10.20431/2349-0365.0505005>
- [3] Kim, N., Fischer, A.H., Dyring-Andersen, B., Rosner, B. and Okoye, G.A. (2017) Research. Techniques Made Simple: Choosing Appropriate Statistical Methods for Clinical Research. *Journal of Investigative Dermatology*, **137**, e173-e178.
<https://doi.org/10.1016/j.jid.2017.08.007>
- [4] Siegel, S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, Singapore.
- [5] Singh, P. (n.d.) 2×2 Contingency Table : Fisher's Exact Test.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4965cdf99774802ec26dcfbf754a924ab874ba23>
- [6] Pembury Smith, M.Q. R. and Ruxton, G.D. (2020) Effective Use of the McNemar Test. *Behavioral Ecology and Sociobiology*, **74**, Article No: 133.
<https://doi.org/10.1007/s00265-020-02916-y>