ISSN 1980-5411 (On-line version)
Production, 35, e20240076, 2025
DOI: 10.1590/0103-6513.20240076

Research Article

Normal nonparametric test for small samples of categorized variables

Jose Luiz Contador[a]*, Edson Luiz França Senne[a], Jose Celso Contador[b]

[a]Universidade Estadual Paulista, Guaratinguetá, SP, Brasil

[b]Universidade Paulista, São Paulo, SP, Brasil

*jose.contador@unesp.br

Abstract

Paper aims: Introduce a new statistical test to verify whether two small samples of variable classified into multiple categories are drawn from the same population. This problem can be represented by a contingency table of order (m x 2).

Originality: We do not have adequate asymptotic texts to treat this issue in all instantiation of the problem, and exact methods require substantial computational effort and specialized algorithms. The proposed test covers this gap.

Research method: It can be classified within design science research. The result, as well as the research process, meets the guidelines of that research method.

Main findings: Computational experiments show that the proposed test has similar effectiveness to the exact test, even when dealing with sparse data contingency tables and small values of m. Furthermore, examples show that it can work well in cases where the chi-square test, its numerous variations, and even in situation where the more recently developed methods fail.

Implications for theory and practice: This type of decision problem has received significant attention in the literature because it represents many real-life situations. The test proposed is as useful for small samples as Chi-square is for larger samples.

Keywords

Nonparametric tests. Small samples. Nominal variables. Permutation tests. Computer simulation.

How to cite this article: Contador, J. L., Senne, E. L. F., & Contador, J. C. (2025). Normal nonparametric test for small samples of categorized variables. *Production*, *35*, e20240076. https://doi.org/10.1590/0103-6513.20240076.

Received: July 6, 2024; Accepted: June 2, 2025.

Financial Support
None

Conflict of Interest
None

Ethical Statement
This research is eminently theoretical and did not involve people in data collection

Editor(s)
Adriana Leiras

1. Introduction

This paper presents a new statistical test designed for decision problems that can be represented by a contingency table of order (m x 2), particularly applicable to scenarios with small sample sizes and sparse data. The objective is to ascertain whether the data in the two columns belong to the same population (null hypothesis H_0).

Sparse data contingency tables have received considerable attention from the literature given their capacity to represent real-life issues and the challenge they pose concerning the null hypothesis, especially when it comes to large values of categories m. However, few studies have been published addressing small samples.

The practical relevance of the subject discussed can be assessed by the extensive applicability that the chi-square test has in solving real-life problems. The method presented in this article allows addressing these types of problems, where we have small samples with sparse or non-sparse data, cases in which, as is known, the chi-square test does not work well.

Many issues in biology, medicine, the social and human sciences, as well as in business and industry, present the characteristics that require the application of the test proposed here. Some examples of problems directly related to social and industrial engineering are presented below (see Contador & Senne, 2016):

– Determining whether two different types of employees (machine operators and office workers, for example) in small companies (with few workers) have similar motivations in order to develop a single incentives program (or include all workers in a single program);

– Determining, through a small sample of companies from different sectors (e.g., manufacturing and services) whether these companies value the same characteristics in their executives to standardize human development programs;

– Determining whether executives (few in number) from different business units of a corporation have similar managerial capacity;

– Determining whether two different production processes, by analyzing few parts, create products with similar levels of quality for different characteristics (size, finishing, etc.).

To illustrate the proposed model, we can examine a specific issue known as the 'strategies comparison problem', which motivated the creation of the test here proposed. This problem arises from empirical research conducted in the development of the Fields and Weapons of Competition model - *FWC* (Contador, 2008; Contador et al., 2023).

The objective of this problem is to verify whether the business strategy adopted by a company is a determining factor of its competitiveness. In his research, Contador (2008) collected a small sample of companies and divided them into two groups using an appropriate criterion. One of the groups was formed by the most competitive companies and the other by the least competitive. The objective was to verify whether both groups adopt similar business strategies (null hypothesis H_0). This problem can be represented by a contingency table of order (m x 2) with a small sample.

A business strategy is represented by the Field of the competition, the imaginary locus of dispute in a market between products or companies for customers' preferences, in which the company seeks to achieve and maintain competitive advantage. Each field of the competition represents an attribute of the product or company that the customer recognizes and values (Contador, 2008). According to the *FWC* model, companies focus their main competitive strategy on one of the 14 fields of competition identified by Contador (2008).

Out of curiosity, in the 19 empirical studies that validated the CAC model, involving 238 companies from different industrial and service sectors (Contador, 2008), it was found that there is no significant difference between the set of competition fields adopted by the most competitive companies and by the least competitive ones. This means that both groups of companies have equal perception of customer preferences and, therefore, the choice of business strategy does not explain why one company is more competitive than another.

In fact, what determines the company's competitiveness is the correct alignment of its essential competence (core competence, according to Hamel & Prahalad, 1995) with the field of competition chosen by the company to compete. The author of the *FWC* model has demonstrated this fact in its studies and constitutes his thesis: "For the company to be competitive there is no condition more relevant than having high performance only in those few weapons that give it competitive advantage in the competition fields chosen for each product/market pair".

Weapon of the competition is an internal company resource capable of conquering and/or maintaining competitive advantage. A company uses more than 100 weapons; around 40 or 50 are considered competition weapons and around 1/3 of them are relevant to compete in a given field of competition. If used with high efficiency the field of competition becomes visible to the customer. This means the correct alignment between weapons and fields of competition. The studies have shown that this property explains 80% of the phenomenon of competitiveness (Contador, 2008).

To better understand the strategies comparison problem, let us examine the data in Table 1. This table, extracted from one of the studies conducted by the author of the *FWC* model, involves 21 companies (E1 to E21, as listed in Table 1). These companies were categorized into two groups based on their degree of competitiveness (*DC*), determined by the variation in income over a specific period (five years).

The respective companies declared the main fields of competition, identified by letters A to F. Thus, the main strategies of both groups of companies can be represented in Table 2.

Therefore, the null hypothesis H_0 considers that the lists of strategies C_1 and C_2 belong to the same population.

This type of test may be done by determining whether the sets of values f_i and g_i (see Table 3) can be considered to belong to the same population, where f_i and g_i are the distributions of the frequencies with which the strategies $i = 1, 2, ..., m$ appear in Group I and Group II of companies, respectively, so that $\sum_{i=1}^{m} f_i = n_1$ and $\sum_{i=1}^{m} g_i = n_2$. In Table 2, f_i and g_i assume the values expressed in Table 3. Therefore, we want to find out if the two groups choose the classes in a similar way, that is, if variables f_i and g_i can be considered as belonging to the same population.

There are two approaches to address this problem: 1) employing asymptotic tests, derived from the chi-square test, or 2) using exact tests, originating from Fisher's exact test. Fisher's exact test is applicable specifically to (2 x 2) tables.

The chi-square and its variations may fail in situations like that. By using simulation data, Yang et al. (2015) found that when 50% of cells in a contingency table have an expected count of less than five, or when there is a zero expected count in any cell, the p-values from likelihood ratio tests exhibit a relative error greater than 100%.

Table 1. Main field of competition and degree of competitiveness of the enterprises.

Group I: Most competitive companies				Group II: Least competitive companies			
Main field of competition (*FC*)				Main field of competition (*FC*)			
Ent	Denomination	Code	DC_i	Ent	Denomination	Code	DC_i
E10	Product and brand image	*A*	1.51	E05	Variety of models	*D*	0.82
E13	Product delivery deadline	*B*	1.43	E11	After-sales service	*C*	0.80
E17	After-sales service	*C*	1.39	E06	Product and brand image	*A*	0.79
E19	After-sales service	*C*	1.32	E12	Product and brand image	*A*	0.79
E21	Variety of models	*D*	1.25	E04	Product and brand image	*A*	0.69
E02	Product and brand image	*A*	1.19	E14	Presales service	*F*	0.62
E08	Product project	*E*	1.16	E16	Product project	*E*	0.54
E03	After-sales service	*C*	1.14	E07	Product and brand image	*A*	0.47
E13	Product project	*E*	1.11	E09	Presales service	*F*	0.38
E01	Variety of models	*D*	1.07	E20	Product project	*E*	0.30
				E18	After-sales service	*C*	0.25

Caption: *FC* = Main field of competition ;Ent = Enterprise; *DC* = Degree of competitiveness
Source: Contador (2008).

Table 2. Strategies adopted by the two groups of firms.

Groups of firms	Strategies										
More competitive (C_1)	*A*	*A*	*B*	*C*	*C*	*C*	*D*	*D*	*E*	*E*	
Less competitive (C_2)	*A*	*A*	*A*	*A*	*C*	*C*	*D*	*E*	*E*	*F*	*F*

Table 3. Frequencies of strategies (*FC*) for the groups of companies.

FC	*i*	f_i	g_i
A	1	2	4
B	2	1	0
C	3	3	2
D	4	2	1
E	5	2	2
F	6	0	2

The classic alternative offered by the literature when there are tables larger than (2 x 2) is given by Freeman & Halton (1951). This alternative, however, requires specific algorithms that generally demand a great computational effort, such as the StatXact software (StatXact, 2008).

Another alternative was offered by Hothorn et al. (2008), which presented the "coin" package for conditional inference. It is the computational counterpart to the theoretical structure presented by Strasser & Weber (1999). However, although more comprehensive than StatXact, it also demands the application of a special algorithm.

Although several authors have suggested alternatives to reduce the effort required to apply exact tests, adopting an appropriated asymptotic tests would still preferable because the exact methods demand special algorithms, which are not always readily available to the statistics user.

In view of this fact, authors have adopted different strategies, such as: a) manipulating the data contained in the cells or in the very chi-square function (Q); or, b) creating new statistics that can converge asymptotically to some known distribution, such as the normal or the chi-square (see section 2.3 and 2.4). Throughout the text, we will see that these arrangements do not work for the strategies comparison problem and, even in cases where the chi-square test fails, the proposed test works well.

Two more recently developed tests deserve our attention, presented in the sections 2.5 and 2.6, respectively: Zelterman (1987) and Plunkett & Park (2019). However, these statistics do not work well for small samples or may fail in several possible instantiations of the strategy problem, as will be seen.

We found a single article in the literature that present nonparametric asymptotic tests for small samples of nominal variables (Contador & Senne, 2016). However, this method only applies to some instantiations of the problem in focus, unlike the test proposed here (see Conclusion section).

In summary, based on the literature review, it seems that there is currently no asymptotic test available that effectively addresses the challenges posed by small sample sizes and sparse data. The test proposed here fills this gap.

It addresses the problem in question by constructing a new statistic. The decision variable is given by the sum of m independent unimodal variables, leading it to converge toward rapidly a normal probability distribution as the number of categories m increases. In the strategy problem, m is the number of different strategies mentioned by the enterprises, that is, the numbers of the lines in Table 3.

The primary finding of this study indicates that the proposed test exhibits effectiveness similar to the exact test. Furthermore, it performs well in scenarios where the chi-square test fails, such as in cases of small samples and sparse data with significant imbalance. Furthermore, it is suitable in situations where other tests, including Zelterman (1987) and Plunkett & Park (2019), do not work and effectively replaces methods that require special procedures, as suggested by Freeman & Halton (1951).

The article is structured into six sections. The second section presents the behavior of nonparametric statistics in addressing the problem at hand. The third section introduces the proposed test and illustrates its application. In the fourth section we present the architecture used in the construction of the simulation process to evaluate the consistency of the proposed test by comparing it to the exact test. The fifth section presents the results of the evaluation of the proposed test effectiveness. The last section presents de conclusions.

2. Nonparametric statistics and the problem of small samples of categorized variables

In this section, we outline the challenges associated with nonparametric statistics and provide a literature review on nonparametric tests of categorized variables, highlighting their limitations in addressing the specific problem (small samples) presented in this study.

The nonparametric tests seek to determine, from the sample data, the ρ true probability value (tail value) that will lead to the decision to reject or not the null hypothesis.

The value of ρ can be evaluated or calculated (p-value, as we know it) in two ways, using the notation adopted by Contador & Senne (2016):

a) Through the expression $\rho = P(X \geq x_{cal})$, where X represents a distribution of known probability and x_{cal} is a value calculated from a function of the sample data, so that $x_{cal} \in X$; or

b) Through the expression $\rho = \sum_{i=1}^{r} p_i$, where p_i, for i=1, is the probability of occurring that configuration of values shown by the sample, and p_i, for i = 2, ..., r, are the probabilities of any of the other (r – 1) possible configurations occurring more extremely than in the original sample. For better understanding, see section 2.1

These two ways of determining the p-value divide the nonparametric tests in two classes: *approximate tests* (or asymptotic), when ρ is determined through (a), described above, and *exact tests*, when ρ is calculated through (b).

The exact tests can be addressed through the permutations theory, initially introduced by Fisher (1970) for contingency tables of dimension (2 x 2). Subsequently, this approach was extended to larger tables (see Freeman & Halton, 1951 and Hothorn et al., 2008).

Asymptotic tests require a large enough sample size to ensure confidence in the p-value obtained. When we have small samples, we should choose the exact tests that determine the true values of p_i, and therefore of ρ.

In the mid-twentieth century, nonparametric methods applied to problems with ordinal variables received great impulse from Wilcoxon (1945), who presented a test based on the sum of the posts of two samples in order to verify if they were extracted from the same population. Later on, Mann and Whitney (1974) developed a more appropriate procedure, which originated the Wilcoxon-Mann-Whitney test.

Other important initial studies in nonparametric statistics, which also address ordinal variables, can be found in the following references: Chernoff & Savage (1958), Friedman (1937), Kruskal & Wallis (1952), Smirnov (1939) and Wald & Wolfowitz (1940).

These studies gave rise to the following nonparametric tests considered classical: sign test; Wilcoxon sign test; test of the Wilcoxon-Mann-Whitney station sum; chi-square test; median test; and t-test for paired samples.

In recent years, parametric statistics has been applied in control charting techniques in the cases where there is not enough information to justify the assumption of a specific form for the underlying process distribution. To this end, nonparametric or distribution-free control charts have been proposed. Information on the subject can be obtained at Chakraborti & Graham (2019).

For an asymptotic statistical test to work properly for the problem in question, the respective test variable x_{cal}, calculated from the data of the two samples and used to determine ρ =P[$X \geq x_{cal}$], must possess three properties: a) consider the amplitude of the difference observed in each pair of values related to each class of the random variable; b) allow to accumulate the differences in opposite directions observed in distinct classes (to prevent one to cancel the other), that is, to calculate $|f_i - g_i|$; and c) adjust to a known probability distribution X.

The only test among the mentioned ones that presents the first two properties is the chi-square. However, in order to meet the third property, at least 80% of the cells must have a frequency higher than 5, and no cell can have a frequency lower than 1 (Siegel & Castellan Junior, 2006). Other tests were suggested later, like Campbell (1976), Zelterman (1987) and Plunkett & Park (2019).

2.1. The Fisher's exact test

Tables 4a-c are examples of the application of Fisher's exact test to tables with a (2 x 2) dimension, as presented in Contador & Senne (2016). In this example, Group I corresponds to the male sex, while Group II corresponds to the female sex.

In the upper line of each of these tables are the frequencies $n_{1,j}$, j=1, 2, of people with a height of 1.80 m or taller. In the lower line, there are the frequencies $n_{2,j}$, j=1, 2, of people who are under 1.80 m tall. These data were obtained from a sample of eight men and nine women. Based on this small sample, the idea is to gauge whether men are taller than women are.

Consider that the H_0 hypothesis establishes equality of height, and the alternative hypothesis, H_1, establishes that men are taller than women are. To apply Fisher's exact test to this problem, the value of $\rho = \sum_{i=1}^{r} p_i$ is determined, where p_i is the probability of an equal or more extreme situation occurring (in the sense of Hypothesis H_1) than that of Table 4a, maintaining the total fixed marginal values, $n_{i,o} = \sum_{j=1}^{2} n_{i,j}$ e $n_{o,j} = \sum_{i=1}^{2} n_{i,j}$.

Observe that the sample included six men ($n_{1,1}$) who were taller than 1.80 m and two ($n_{2,1}$) who were shorter. As the test is unilateral, (due to the alternative hypothesis H_1), there are two other more extreme situations than that of Table 4a with fixed marginal values, which are represented in Tables 4b and 4c.

Table 4. Data to exemplify Fisher's Exact Test.

Groups			Groups			Groups		
I	II		I	II		I	II	
6	3	9	7	2	9	8	1	9
2	6	8	1	7	8	0	8	8
8	9	17	8	9	17	8	9	17
	(a)			(b)			(c)	

Source: Contador & Senne (2016).

The exact probability of observing a particular set of frequencies in a (2 x 2) table, when the marginal totals are considered fixed, is given by the hypergeometric distribution, resulting in ρ = 0.109, obtained from the sum of $p_{(a)}$, $p_{(b)}$ and $p_{(c)}$, given by:

$$P = \frac{\prod_i n_{i,o}! \prod_j n_{o,j}!}{n! \prod_{i,j} n_{i,j}} \tag{1}$$

which provides $p_{(a)} = 0.0968$; $p_{(b)} = 0.0012$ and $p_{(c)} = 0.0004$, resulting ρ > 0.05, which indicates that we can't reject H_0 at a significance level α=0.05.

2.2. The Fisher's exact test extension for larger tables

Consider the example from Table 5, whose data refer to the number of executives who belong to four business units of a large corporation and have been given high, average, and low evaluations in an executive promotion program. Based on this small sample, is it possible to conclude that Business Unit *A* has the most capable executives (alternative hypothesis H_1)?

If the chi-square tests were applied, the constructed statistic would have $(l-1)\times(c-1)$ = 6 degrees of freedom and would supply χ^2 = 11.555. As P($\chi_6^2 > 11.555$) = 0.0726, from which it is concluded that there is no statistical evidence that Business Unit *A* has more capable executives.

To apply the exact test, all the possible tables are generated from the configuration of the sample data, maintaining fixed marginal values. The tables that originate values of $\chi^2 \geq 11.555$ represent more extreme situations than the original sample and thus contribute with their respective values of *p* to compose the value of ρ.

For instance, Tables 6a and 6b are two possible arrangements obtained from Table 5. The first is $\chi^2 = 14.676$, and should be considered a more extreme situation than the original model. Thus, its respective value of *p* contributes to the determination of ρ. Meanwhile, Table 6b provides $\chi^2 = 9.778$, and its corresponding value of *p* does not contribute to the calculation of ρ.

The calculation of probability *p* of a particular set of frequencies for a table with *l* rows and *c* columns, according Freeman & Halton (1951), is made using a generalization of Equation 1, for $i = 1,2,\ldots,l$ and $j = 1,2,\ldots,c$

When applying the exact test to tables of dimension $l \times c$, all the possible tables from the originating data of the sample must be represented. The representation of these tables generally requires considerable computational effort.

This type of problem can be solved using software such as the StatXact (2008). For this particular case, this software arrives at ρ = 0.0398, which, contradicting the result of the chi-square test.

Several authors have suggested alternatives to reduce the effort required to apply exact tests. Mehta & Patel (1983) presented a Network Algorithm for performing the Exact Test in (*r* x *c*) contingency tables, which

Table 5. Result of the evaluation of executives.

Evaluation Level	Business Units				
	A	B	C	D	Total
High	5	2	2	0	9
Average	0	1	0	1	2
Low	0	2	3	4	9
Total	5	5	5	5	20

Source: Contador & Senne (2016).

Table 6. Two permutations of the results of the evaluation of the executives.

Business Units					Business Units				
A	B	C	D	Total	A	B	C	D	Total
5	2	2	0	9	4	3	2	0	9
0	0	0	2	2	1	0	0	1	2
0	3	3	3	9	0	2	3	4	9
5	5	5	5	20	5	5	5	5	20
		(a)					(b)		

was then considered the best algorithm to deal with this problem, according to Hirji & Johnson (1996). Some modifications in the Network Algorithm for (2 × *c*) tables were proposed in Requena & Ciudad (2006), which drastically reduced computation time. In certain cases, this reduction exceeds 99.5% compared to StatXact, as reported by the authors.

Another way of treating sparse contingency tables is by the Markov chain Monte Carlo exact tests, which are a powerful tool. Diaconis & Sturmfels (1998) proposed an algebraic algorithm to construct a connected chain over the two-way contingency tables with fixed sufficient statistics and an arbitrary configuration of structural zero cells. Aoki & Takemura (2005) observed that their algorithm did not seem to provide a satisfactory answer because the Markov basis produced by the algorithm often contains many redundant elements and is hard to interpret. Thus, they derived an explicit characterization of a minimal Markov basis, proved its uniqueness, and presented an algorithm for obtaining the unique minimal basis.

2.3. The manipulation of the *Q*-function

As an example of manipulation of the *Q*-function we can cite Lawal (1984), who suggested the modification $Q^* = (1-\beta/N)Q$, where $\beta = 2/3$ e $\beta = 3/2$, for tests with the significance level equal to 0.05 and 0.01, respectively, and N is the total number of sample elements, that is, $N = \sum_i n_{i,o} = \sum_j n_{o,j}$.

However, the use of Q^* has been recommended under restrictive conditions, according Lin et al. (2015): (*i*) The smallest cell expectation e should satisfy the constraint $e \geq s.k^{(-3/2)}$, where s is the number of cells having expectations (under H_0) less than 3 and $k=(l-1).(c-1)$; (*ii*) The dimensions of the table should satisfy $2l > c \geq l > 2$, which clearly does not correspond to the structure of the strategy problem.

Another typical recommendation in the conventional methods is adding a small constant to every cell of the observed table (Subbiah et al., 2008). However, this procedure disfigures the strategies comparison problem and cannot be used in this case.

2.4. Creating new statistical of tests

Strasser & Weber (1999) introduced the likelihood ratio chi-square test statistic represented by Equation 2.

$$Q^l = 2.\sum_{i=1}^{l}\sum_{j=1}^{c} x_{i,j}\, ln(\pi_{i,j} / \pi_{i,j}^0) \qquad (2)$$

where $\pi_{i,j} = x_{i,j}/N$, $x_{i,j}$ is given by the vector of frequencies $(f_i, g_i; x_{i,o} = f_i, x_{o,j} = g_i)$, $N = (n_1 + n_2)$ and $\pi_{i,j}^0$ is the value of $\pi_{i,j}$ when H_0 is true.

However, according to Lin et al. (2015), the chi-square approximation of this function is usually poor when $N/(l.c) < 5$, where l and c are the dimension of the table, rows, and columns numbers. In the strategy problem, we usually have $l = m > 3$, and $c = 2$. In this problem, if the original table presents null cells ($f_i = 0$ and/or $g_i = 0$), it is necessary to keep the original value.

2.5. The Zelterman estimate

Zelterman (1987) proposed the following test statistic D^2 for large contingency tables with sparse values that, according to the author, are close to the normal distribution:

$$D^2 = \sum_{ij} [(x_{ij} - \theta_{ij})^2 - x_{ij}] / \theta_{ij} \qquad (3)$$

where $i = 1,2,\ldots,I$ and $j = i,2,\ldots,J$ are the indexes of lines and columns of the frequencies table $x_{ij} = (f_i, g_i)$,

$\theta_{ij} = \left(\sum_i x_{ij}\right).\dfrac{\left(\sum_j x_{ij}\right)}{\left(\sum_{i,j} x_{ij}\right)}$, $N = \sum_{i,j} x_{ij}$, with D^2 mean and variance given by

$Mean(D^2) = \left(\dfrac{N}{N-1}\right).(I-1).(J-1) - I.J$ and

$$Var\left(D^2\right)=\left(\frac{2N}{N-3}\right).(\gamma-\sigma).(\mu-\sigma)-4.\sigma.\tau/(N-1),$$

being

$$\gamma=(I-1).(N-I)/(N-1);\ \mu=(J-1).(N-J)/(N-1),$$

$$\sigma=\left(N.S-I^2\right)/(N-2),\ \tau=\left(N.T-J\ \right)/(N-2)$$

$$S=\sum_i\left(\sum_j (x_{ij})^{-1}\right) \text{ and } T=\sum_j\left(\sum_i (x_{ij})^{-1}\right)$$

Kim et al. (2009) compare the power of Zelterman's statistic with the chi-square by simulation. Although these authors strongly recommend the use of Zelterman's estimate when the given contingency table is very sparse, this estimator may fail in some situations (see section 5.2)

2.6. Plunkett and Park test

With the aim of testing the hypothesis of equality between two frequency vectors $\{(f_i);(g_i)\}$, i = 1, ..., m, j = 1, 2, under sparse data, Plunkett & Park (2019) propose using the T statistic expressed by Equation 4, and demonstrate that it has distribution N(0, 1) for large values of m.

$$T=\sum_{i=1}^{m}F(f_i,g_i)\Big/\sigma \tag{4}$$

$$F(f_i,g_i)=\left(\frac{f_i}{n_1}+\frac{g_i}{n_2}\right)^2-\frac{f_i}{n_1^2}-\frac{g_i}{n_1^2}$$

$$\sigma^2=\sum_{i=1}^{m}\left(p_i^2-\frac{p_i}{n_1}\right)+\sum_{i=1}^{m}\left(q_i^2-\frac{q_i}{n_2}\right)+\frac{4}{n_1 n_2}\sum_{i=1}^{m}p_i.q_i$$

$$p_i=\frac{f_i}{n_1} \text{ e } q_i=\frac{g_i}{n_2}$$

Although this test is very useful, as it avoids the use of exact methods and possible Chi-square failures, it can lead to wrong decisions for the strategy problem (see section 5.2). This occurs because the T statistic does not converge to the normal probability distribution for small and moderate values of m.

3. The *N*-Normal nonparametric test for small samples of categorized variables

In this section, a new test is presented to verify whether two small samples of categorized variables can be considered belonging to the same population. Test statistic, Z_{cal}, used to determine the value of ρ(N), has a probability distribution that approaches the normal as the number of m classes increases.

Let $\boldsymbol{F}$ = $\{f_i\}$ and $\boldsymbol{G}$ = $\{g_i\}$ be the frequency distribution of each class i = 1, ..., m of the variable as they appear in each of the samples A_1 and A_2, respectively, such that (f_i+g_i) > 0. Let also $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ be the size of the respective samples, that is, $\sum_{i=1}^{m}f_i = n_1$ and $\sum_{i=1}^{m}g_i = n_2$, where $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are moderate values in relation to m.

Notice that F and G represent multinomial probability distributions; therefore, each element f_i and g_i, i = 1, 2, ..., m has a binomial distribution.

Consider the probability distributions P = $\left[p_i=f_i/n_1\right]$ and Q = $\left[q_i=g_i/n_2\right]$. If H_0 is true, the mean a_i and variance b_i of each element p_i or q_i can be determined by the expressions $a_i=(f_i+g_i)/(n_1+n_2)$ and $b_i=a_i(1-a_i)/(n_1+n_2)^{1/2}$, respectively. So, if H_0 is true, variable $T=\sum_{i=1}^{m}Y_i$, where $Y_i=(p_i-q_i)$, has a mean equal to zero, for m big enough, and variance equal to b = $\sum_{i=1}^{m}2b_i$. Notice that Y_i results in a unimodal distribution, since p_i and q_i have the same probabilities distribution. Thus, $Z = \left(T/\sqrt{b}\right)$ converges rapidly to the (0, 1) normal distribution as *m* grows.

Consider now the probabilities distribution *R* and *S,* whose variables are given respectively by $r_i = max\left(p_i, q_i\right)$ and $s_i = min\left(p_i, q_i\right)$, for i = 1, 2, ..., *m*. Since the sets *PQ* = { p_i } ∪ { q_i } and *RS* = { r_i } ∪ { s_i }, para i = 1, ..., *m* are identical, each of the variables r_i and s_i keeps the same properties as variables p_i and q_i.

Let be the function $V_i = \left\{v_i = \left(r_i - s_i\right), i = 1, 2, \ ,m\right\}$. This function refutes all possible negative values ($p_i - q_i$) ϵ Y_i to its opposite. Therefore, if H_0 is true, P[Y_i = *a*] = P[Y_i = -*a*]= ½ P[V_i = *a*], for an integer value of *a*. So, P$\left[T \geq y\right]$ = ½ P$\left[W \geq y\right]$], for y≥0, where W = $\sum_{i=1}^{m} v_i$

Therefore, it is possible to test H_0 through the statistic

$$Z_{cal} = \sum_{i=1}^{m} v_i / \sqrt{b} \tag{5}$$

and calculate the tail value for the test *N* by

$$\rho\left(N\right) = 2.\mathrm{P}\left[Z \geq Z_{cal}\right]. \tag{6}$$

with Z_{cal} being N(0, 1)

The original bilateral hypotheses test, H_0: ***F*** = ***G*** against H_1: ***F*** ≠ ***G*** was converted into the unilateral equivalent test H_0: μ(*W*) = 0 against H_1: μ(*W*) > 0. This change was necessary because is not possible to consider the direction in which the difference occurs between each f_i ϵ***F*** and g_i ϵ***G***, i = 1, ..., *m*.

To exemplify the application of the test proposed, let's consider the strategies' problem mentioned in the Introduction. Table 1 illustrates a real study. It shows 21 firms classified in each of the two groups and their competitive strategy. Six strategies were mentioned by the companies, which are indicated by letters *A* to *F*.

The problem is then to test whether sets C_1 = {*A*, *A*, *B*, *C*, *C*, *C*, *D*, *D*, *E*, *E*} and C_2 = {*A*, *A*, *A*, *A*, *C*, *C*, *D*, *E*, *E*, *F*, *F*} of strategies mentioned by the more and less competitive companies, respectively, can be considered as originating from the same population (null hypothesis, H_0).

Table 7 shows the application of the proposed test to strategies' lists C_1 and C_2. In the last row of column v_i we get the variable *W* = 0.691, and the last row of the last column shows the value of $b = \sum_{i=1}^{m} 2b_i$ = 0.346. From this, we can determine Z_{cal} = 0.691/ $\sqrt{0.346}$ = 1.174, which provides ρ(*N*) = 2.Pr[*Z* ≥ 1.174] = 0.240. Thus, we conclude that H_0 cannot be rejected, and we accept that both groups of firms choose similar sets of business strategies.

The main properties of the proposed test, which validate its use in problems with categorized variables in general, as well as in the strategy problem described here, are:

a) The variable Y_i results in a unimodal distribution, since p_i and q_i have the same probabilities distribution. Thus, $T = \sum_{i=1}^{m} Y_i$, converges rapidly to the normal distribution as *m* grows;
b) The proposed test presents a performance similar to the Exact test according to computational simulation carried out to estimate the power of both tests, as shown in the next section;

Table 7. Application of the proposed test to the strategies' lists C_1 and C_2.

Strategies	i	f_i	g_i	p_i	q_i	a_i	v_i	$2b_i$
A	1	2	4	.200	.364	.286	.164	.089
B	2	1	0	.100	.000	.048	.100	.020
C	3	3	2	.300	.182	.238	.118	.079
D	4	2	1	.200	.091	.143	.109	.053
E	5	2	2	.200	.182	.190	.018	.067
F	6	0	2	.000	.182	.095	.182	.038
Total		10	11	1.000	1.000	1.000	.691	.346

$p_i = \frac{f_i}{\sum_{i=1}^{m} f_i}; q_i = \frac{g_i}{\sum_{i=1}^{m} g_i}; \quad a_i = \frac{f_i + g_i}{\sum_{i=1}^{m} f_i + \sum_{i=1}^{m} g_i}; \quad v_i = \left|p_i - q_i\right|;$

$2b_i = 2.a_i\,(1 - a_i)\,/\,(\sum_{i=1}^{m} f_i + \sum_{i=1}^{m} g_i)^{1/2}$

c) As we know, the chi square test does not work satisfactorily in problems of small sample and sparse data. An example presented in section 5.2 confirms and shows that the proposed test leads to a correct decision in these cases;

d) The test proposed by Zelterman (1987) or by Plunkett & Park (2019) may fail for different instantiations of the problem, or in case of small samples, because the variables of both tests do not converge to the normal distribution in this situation, as will be seen later. On the other hand, the test proposed here responds well in these situations, as shown by some examples presented in section.5.2

4. Method to evaluation the effectiveness of the proposed test

Regarding the research method, we can classify it within design science research. This typology aims to develop standards, strategies and actions to improve results available in the literature, find optimal solutions for new problems or even compare the performance of strategies regarding the same problem (Bertrand & Fransoo, 2002).

The evaluation of consistency (or effectiveness) of the proposed test was carried out through its power curve, which provides the probability of acceptance (*Pa)* of the null hypothesis (H_0), according to the level of similarity of the two samples. It was possible to extract the value of risks α and β from this curve, constructed by computer simulation. In this section we will present the architecture used in the construction of this simulation process, adopting the same procedure used by Contador & Senne (2016).

The power curve was built by varying the level of similarity between the two samples, which was defined by the parameter named *degree of symmetry* (*DS*) between the distributions of samples A_1 and A_2, as in Equation 7

$$DS = \left(\sum_{i=1}^{m} \left|p_i - q_i\right|\right) / 2 \qquad (7)$$

where p_i and q_i are the probabilities of the categorized variable originated from samples A_1 and A_2 for every $i = \{1, 2, ..., m\}$.
When $p_i = q_i$ for every *i*, Equation 7 provides $DS = 0$, and the samples obtained by simulation come from the same population. On the other hand, if $p_i = 0$ when $q_i \neq 0$, for every *i*, then $DS = 1$, which gives rise to configurations with samples belonging to statistically different populations. Therefore, *DS* is defined in the interval [0, 1].

Appropriate values were defined for $p_i\, e\, q_i$ in order to get samples from populations with the following degrees of symmetry: 0.0, 0.2, 0.4, 0.6, and 0.8. Table 8 displays values of $p_i \textit{ and } q_i$ that originate samples drawn from populations with different values of *DS* for $m = 6$.

Problems were generated for the following five cases, defined by the sets of values of (m, n_1, n_2): (3, 7, 7), (4, 8, 8), (5, 10, 10), (6, 12, 12) and (7, 14, 14). For each of these five cases and for each of the five *DS* values previously mentioned, we determined the probability of acceptance P_a according to the exact and normal tests.

We generated 100 problems for each combination of (m, n_1, n_2) and *DS*, originating 2500 problems. For both test, and a given configuration (m, n_1, n_2), and for a given *GS* value, P_a could then be identified by directly counting the number of problems in which H_0 was accepted.

We adopted the significance level α=0.05 for all tests. Thus, the acceptance of H_0 occurred whenever the statistical test yielded value ρ= P[$Z > Z_{cal}$]> 0,05, where *Z* is the test variable and $Z_{cal} = W / \sqrt{2\sum_{i=1}^{m} b_i}$ is the value of the test statistic (Equation 5), being *Z* normal (0, 1).

Table 8. Values of p_i and q_i to get samples with different *GS* values (m=6).

Class	*DS*=0.0		*DS*=0.2		*DS*=0.4		*DS*=0.6		*DS*=0.8	
(i)	p_i	q_i	p_i	q_i	p_i	q_i	p_i	q_i	p_i	q_i
1	1/6	1/6	1/5	2/15	7/30	1/10	4/15	1/15	3/10	1/30
2	1/6	1/6	1/5	2/15	7/30	1/10	4/15	1/15	3/10	1/30
3	1/6	1/6	1/5	2/15	7/30	1/10	4/15	1/15	3/10	1/30
4	1/6	1/6	2/15	1/5	1/10	7/30	1/15	4/15	1/30	3/10
5	1/6	1/6	2/15	1/5	1/10	7/30	1/15	4/15	1/30	3/10
6	1/6	1/6	2/15	1/5	1/10	7/30	1/15	4/15	1/30	3/10

The configuration (or instantiation) of each problem, that is, values of f_i and , i= 1, ... m, for both samples, for a given value of m, providing samples around the values of DS was obtained by the Monte Carlo method, described below:

Step 1. We established the correlation (see Table 9) between the rectangular random number (NA) defined in the interval [0, 1] and the class of variable i, where a_0 = 0, a_m = 1 and, for i=2, ..., m-1, $a_i = (a_{i-1} + p_i)$ for sample A_1, and $a_i = (a_{i-1} + q_i)$ for sample A_2, and q_i mentioned in Equation 7.

Step 2. n_1 rectangular random numbers (NA) were generated in the interval [0, 1] for the first sample and other n_2 random numbers for the second sample. The f_i e g_i values for the first and second samples were determined by counting random numbers obtained in the interval corresponding to the class of variable i.

To illustrate, suppose you want to obtain the value of f_1 for a sample with DS=0.6, n_1 = n_2 =12 e m = 6. To do this, you must simulate twelve rectangular random values in the interval [0, 1] and count how many of them fall in the interval [0, 4/15). To obtain the values of g_1 proceed in the same way and count how many values were now generated in the interval [0, 1/15).

5. Results

In this section we will present the results of the evaluation of the proposed method effectiveness obtained from the simulation process discussed in the previous section. A comparison of the method with other tests provided in the literature is too presented

5.1. The effectiveness of the proposed method

Table 10 shows the results from applying the proposed test (N-Normal) and the Exact test. P_a values are expressed as a percentage because they correspond directly to the number of times that hypothesis H_0 was

Table 9. Correlation between NA and the i variable class.

NA at interval	Variable class (i)
$[a_0, a_1)$	1
$[a_1, a_2)$	2
$[a_2, a_3)$	3
	Etc.
$[a_{m-1}, a_m]$	m

Table 10. Results of computational tests.

	Probability of acceptance of H_0 (P_a)					Risks (%)	
	DS=0.0	DS=0.2	DS=0.4	DS=0.6	DS=0.8	α	β
Test			Case 1: Results for (m = 3, n_1 = 7, n_2 = 7)				
Exact	98	94	80	24	0	2	50
Normal	96	85	75	23	5	4	47
Test			Case 2: Results for (m = 4, n_1 = 8, n_2 = 8)				
Exact	97	94	78	51	12	3	59
Normal	96	90	62	36	6	4	49
Test			Case 3: Results for (m = 5, n_1 = 10, n_2 = 10)				
Exact	96	92	78	34	3	4	52
Normal	95	89	71	23	2	5	46
Test			Case 4: Results for (m = 6, n_1 = 12, n_2 = 12)				
Exact	90	94	71	28	6	10	50
Normal	87	85	64	20	4	13	43
Test			Case 5: Results for (m = 7, n_1 = 14, n_2 = 14)				
Exact	90	94	67	32	3	10	49
Normal	90	91	58	25	2	10	44

accepted in the 100 problems tested for each *DS* value. Columns α and β show the values of the risks related to the error types I and II obtained from the tests. The values of α are given by $(1 - P_a)$ for *DS* = 0, and that of β resulting from the arithmetic mean of P_a for the cases where *DS* > 0.

By analyzing the *DS* = 0.0 and *DS* = 0.8 columns of Table 10, we verified that the proposed test showed a performance similar to the Exact test for these two extreme cases of population similarity. Note that even for small dimension tables (*m* = 3 - first block in Table 10), we had a good performance of the proposed test, which may indicate a fast convergence of the test statistic *Z* (*N*-Normal test) to the normal probability distribution.

The P_a values can also be used to identify the number of tests in which it led to the right decision, that is, to accept hypothesis H_0 when it is true (P_a values for *DS* = 0) and to reject it when it is false (sum of the values of [100- P_a] for the cases where *DS*> 0). These values are shown in Table 11.

Table 11 shows that, of the 2.500 problems tested, 1.548 were correctly decided by the proposed test, and 1.436 by the Exact test. By examining columns *DS* = 0 and *DS* > 0, we see that the proposed test gives a somewhat lesser protection compared to the exact test, in relation to Type *I* error, but exceeds it with respect to Type *II* error. This is an interesting result since type II risk cannot be controlled in hypothesis testing.

It is common to use computational simulation to estimate the power of statistical tests (Tanizaki 1997). Plunkett & Park (2019) also use simulation to compare their method with some others proposed in the literature.

Hence, based on the tests carried out, we can accept the hypothesis that the test proposed here is a valid alternative to the exact test, especially if the number of problem (*m)* classes is not small.

5.2. Comparison of the proposed method with other tests

In this section we observe how the proposed test positions itself in relation to traditional chi-square and exact method approaches. In addition we showing its contribution through comparison with the main tests presented in the literature.

In section 2 we showed the two ways of determining the p-value, which divides the nonparametric tests in two traditional approaches: *approximate tests* (or asymptotic) and *exact tests*. The proposed test belongs to the first one, whose most commonly used representative is chi-square. When comparing the proposed test with the chi-square, first of all, it is important to analyze its behavior in situations where this test does not work well.

The example in Table 12 may provide some clues. By applying the Exact test to the data in this table, we obtain, through StatXact (2008), ρ = 0.0013, which shows that the three samples do not come from the same population. The chi-square test, in turn, gives a value of ρ = 0.1342, clearly showing that for small samples with sparse data and strong imbalance, like the case in this example, it does not work well.

To assess the performance of the proposed test for these types of samples, it is important to note that while it was initially designed for two-sample problems, it can also be extended to scenarios with a larger number of samples. This involves applying the test to various combinations of samples, two by two.

Doing this with Table 12 data, we get ρ values of 0.002, 0.133, and 0.001 for the sample combinations *A*/*B*, *A*/*C* and *B*/*C*, respectively. This shows, with a high degree of confidence, that sample *B* comes from a different population than the others, in agreement with the result of the application of the exact test.

We also saw that data manipulation contained in the cells or in the chi-square function *Q* itself (see section 2.3), or creating new statistics that can converge asymptotically to some known distribution, such as the normal or the chi-square (see section 2.4) disfigures the strategy comparison problem and cannot be used in this case. Therefore, they are unreliable options for treating the problem in question.

Table 11. Number of problems with right decision.

Test	*DS*=0	*DS*=0.2	*DS*=0.4	*DS*=0.6	*DS*=0.8	*DS*>0	Total
Exact	471	32	126	331	476	965	1436
Normal	464	60	170	373	481	1084	1548

Table 12. Example of a problem with three samples.

Sample	Values								
A	0	7	0	0	0	0	0	1	1
B	1	1	1	1	1	1	1	0	0
C	0	8	0	0	0	0	0	0	0

Source: StatXact (2003).

To compare with the Zelterman estimate (1987) consider the instantiation $\{(f_i);(g_i)\}$ = {(1, 0, 0, 0, 1, 1, 1, 1, 1)^t; (1, 1, 1, 1, 0, 0, 0, 0, 0)^t } of the strategy problem, which shows a clear distinction between the strategies of the two groups of companies (they are practically opposite). While the Zelterman estimate gives ρ(D^2)=0.817, we have ρ(*N*)= 0.025, which shows that Zelterman does not work well for this instantiation while the *N*-Normal test is consistent, since there is a weak intersection of the set of strategies of both groups of companies.

The Zelterman's estimate presents another problem. Note that if $N=S=I$, $\gamma=\sigma=0$ we have $Var\left(D^2\right)=0$, what happens with the instantiation given by $\{(f_i);(g_i)\}$ = {(0, 0, 0, 0, 1, 1, 1, 1, 1)^t; (1, 1, 1, 1, 0, 0, 0, 0, 0)^t}, an possible occurrence for the strategy problem and others. This instantiation leads to ρ(*N*)= 0.009, which also seems to be consistent, since the intersection of the sets of strategies of both groups of companies is empty

With respect to the Plunkett & Park test (2019), we observed the same phenomenon. The instantiation for the strategy problem mentioned above results *T*=1,14, providing *p*-value = 0,13.

Based on the computational tests conducted and the examples presented, we can assert that the proposed test demonstrates effectiveness comparable to the Exact test. Furthermore, it appears to perform well in scenarios where the chi-square and even the Zelterman estimate or the Plunkett & Park test fail.

6. Conclusions

The test proposed here arose from the idea of verifying whether the differences $d_i=|f_i-g_i|$ are statistically significant, and the critical issue of the proposed method lies in the assumption that the variable $T=\sum_{i=1}^{m}Y_i$ converges to the normal distribution for small values of *m*.

Since f_i e g_i has a binomial probability distribution, d_i also maintains this property. In addition, we know that this distribution quickly approaches the normal one, which justifies the hypothesis assumed about the variable *T*, resulting from the sum of unimodal variables, such as the binomial.

This hypothesis is confirmed by the computational tests carried out, once it shows that, even for the case of *m*=3, the proposed test presents similar performance to the exact test. Additionally, a simulation with 30 cases where $(n_1+n_2)\leq 20$, showed that it is not possible to reject the hypothesis H_0: Z_{cal} fits a normal distribution *N*(0, 1), according to the Kolmogorov-Smirnov, Jarque-Bera, D'Agostino and Shapiro-Wilk tests.

On the other hand, this is not what happens with variables D^2 of Zelterman e *T* of Plunkett & Park. Both are based on a quadratic function of the difference between f_i e g_i, *i*= 1, 2, ..., *m*, whose approximation to the normal occurs for large values of *m*. Both authors demonstrated convergence to the normal distribution, but under this condition.

In fact, submitting both variables D^2 e *T* to the same 30 simulated cases to the Kolmogorov-Smirnov, Jarque-Bera, D'Agostino and Shapiro-Wilk tests, all led to the rejection of adherence to the normal distribution of probabilities, when *m*=3.

Finally, it should be noted an important property of the proposed test. The differences $|f_i-g_i|$, considering the instantiation of the strategy problem given by $\{(f_i);(g_i)\}$ = {(0, 0, 0, 0, 1, 1, 1, 1, 1)^t; (1, 1, 1, 1, 0, 0, 0, 0, 0)^t} and that given by $\{(f_i);(g_i)\}$ = {(2, 2, 2, 2, 1, 1, 1, 1, 1)^t; (1, 1, 1, 1, 2, 2, 2, 2, 2)^t} are equal.

However, from a logical standpoint, a significant distinction exists between them. In the first instantiation, the sets of strategies chosen by both groups of companies have an empty intersection, whereas in the second instantiation, there is a substantial intersection between them. The proposed test effectively captures this distinction, providing ρ(N) = 0.009 for the first instantiation and ρ(N) = 0.260 for the latter.

There are very few articles available in the literature that present nonparametric asymptotic tests for small samples of nominal variables. In fact, in a search process carried out in the Scopus database with the words "Nonparametric" OR "Non-parametric" AND "nominal variables" OR "categorized variables" OR "multiple categories" in the fields "Article title, abstract and keywords", 56 documents appear, since 1985. If we add the expression "small sample" in the search process, the single article appears is Contador & Senne (2016).

These authors presented a new nonparametric asymptotic test for small samples of categorized variables based in the difference *D* between two uniform distributions of probabilities. However, the *D* distribution is not known. To get around this problem, the authors constructed, through simulation, their histogram for some values of (*m*, n_1, n_2) and gets the critical values of the test variable *D* (D_α , α= 0,01 and 0,05). Although it has shown good effectiveness compared to the exact test, the method proposed does not apply for other values of (*m*, n_1, n_2).

Due to the scarcity of tests aimed at the subject of this article, the chi-square with its variations and Zelterman (1987) e Plunkett & Park (2019, although aimed at large samples, were included in this article because we was interest in verifying whether they would respond well in the case of small samples and/or sparse data, which was not the case. The first ones mentioned are classical methods and the other two were more recently developed.

From everything was seen, it seems that the test proposed here becomes a viable alternative of exact tests (perhaps the only one) for decision problems involving two small samples classified into multiple categories, with sparse data or not.

Data availability

Research data is available in the body of the article.

References

Aoki, S., & Takemura, A. (2005). Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *Journal of Statistical Computation and Simulation*, *75*(10), 787-812. http://doi.org/10.1080/00949650410001690079.

Bertrand, J.W.M., & Fransoo, J. (2002). Operations Management Research Methodologies Using Quantitative Modeling. *International Journal of Operations & Production Management*, *22*, 241-264. http://doi.org/10.1108/01443570210414338.

Campbell, B. R. (1976). Partitioning chi-square in contingency tables: a teaching approach. *Communications in Statistics. Theory and Methods*, *6*(6), 553-562. http://doi.org/10.1080/03610927708827513.

Chakraborti, S., & Graham, M. A. (2019). Nonparametric (distribution-free) control charts: an updated overview and some results. *Quality Engineering*, *31*(4), 523-544. http://doi.org/10.1080/08982112.2018.1549330.

Chernoff, H., & Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Annals of Mathematical Statistics*, *29*(4), 972-994. http://doi.org/10.1214/aoms/1177706436.

Contador, J. C. (2008). *Campos e armas da competição: novo modelo de estratégia*. São Paulo: Ed. Saint Paul.

Contador, J. C., Contador, J. L., & Satyro, W. C. (2023). CAC-Redes: a new and quali-quantitative model to increase the competitiveness of companies operating in business networks. *Benchmarking*, *30*(10), 4313-4341. http://doi.org/10.1108/BIJ-03-2022-0204.

Contador, J. L., & Senne, E. L. F. (2016). Testes não paramétricos para pequenas amostras de variáveis não categorizadas: um estudo. *Gestão & Produção*, *23*(3), 588-599. http://doi.org/10.1590/0104-530x357-15.

Diaconis, P., & Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, *26*(1), 363-397. http://doi.org/10.1214/aos/1030563990.

Fisher, R. A. (1970). *Statistical methods for research workers* (14th ed.). Edinburgh: Oliver and Boyd.

Freeman, G. H., & Halton, J. H. (1951). Note on an exact treatment of contingency goodness-of-fit and other problems of significance. *Biometrika*, *38*(1-2), 141-149. http://doi.org/10.1093/biomet/38.1-2.141. PMid:14848119.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*(200), 675-701. http://doi.org/10.1080/01621459.1937.10503522.

Hamel, G., & Prahalad, C. K. (1995). *Competindo pelo Futuro: estratégias inovadoras para obter o controle do seu setor e criar os mercados de amanhã* (10. ed.). Rio de Janeiro: Campus.

Hirji, K. F., & Johnson, T. D. (1996). A comparison of algorithms for exact analysis of unordered 2×k contingency tables. *Computational Statistics & Data Analysis*, *21*(4), 419-429. http://doi.org/10.1016/0167-9473(94)00021-2.

Hothorn, T., Hornik, K., van de Wiel, M. A., & Zeileis, A. (2008). Implementing a class of permutation tests: the coin package. *Journal of Statistical Software*, *28*(8), 1-23. http://doi.org/10.18637/jss.v028.i08.

Kim, S.-H., Choi, H., & Lee, S. (2009). Estimate-based goodness-of-fit test for larges parse multinomial distributions. *Computational Statistics & Data Analysis*, *53*(4), 1122-1131. http://doi.org/10.1016/j.csda.2008.10.011.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583-621. http://doi.org/10.1080/01621459.1952.10483441.

Lawal, H. B. (1984). Percentile values of the $\chi 2$ statistic in small contingency tables. *The Indian Journal of Statistics*, *46*(1), 64-74. Retrieved in 2024, July 6, from https://www.jstor.org/stable/25052326

Lin, J.-J., Chang, C.-H., & Pal, N. (2015). A revisit to contingency table and tests of independence: bootstrap is preferred to chi-square approximations as well as fisher's exact test. *Journal of Biopharmaceutical Statistics*, *25*(3), 438-458. http://doi.org/10.1080/10543406.2014.920851. PMid:24905809.

Mann, H. B., & Whitney, D. R. (1974). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*(1), 50-60. http://doi.org/10.1214/aoms/1177730491.

Mehta, C. R., & Patel, N. R. (1983). A network algorithm for performing Fisher's Exact Test in r×c contingency tables. *Journal of the American Statistical Association*, *78*(382), 427-434. http://doi.org/10.2307/2288652.

Plunkett, A., & Park, J. (2019). Two-sample test for sparse high-dimensional multinomial distributions. *Test*, *28*(3), 804-826. http://doi.org/10.1007/s11749-018-0600-8.

Requena, F., & Ciudad, N. M. (2006). A major improvement to the Network Algorithm for Fisher's Exact Test in (2 x *c*) contingency tables. *Computational Statistics & Data Analysis*, *51*(2), 490-498. http://doi.org/10.1016/j.csda.2005.09.004.

Siegel, S., & Castellan Junior, N. J. (2006). *Estatística não-paramétrica para ciência do comportamento* (2. ed.). Porto Alegre: Artmed.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, *2*(2), 3-14.

StatXact. (2003). *Software for small-sample categorical and nonparametric data: user manual Version 6*. Cambridge: Cytel Software Corporation.

StatXact. (2008). Software for small-sample categorical and nonparametric data. Free Trial. Cambridge: Cytel Software Corporation. Retrieved in 2024, July 6, from http://www.cytel.com/products/statxact/

Strasser, H., & Weber, C. (1999). On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics*, *8*(2), 220-250. http://doi.org/10.57938/ff565ba0-aa64-4fe0-a158-86fd331bee78.

Subbiah, M., Kumar, B. K., & Srinivasan, M. R. (2008). Bayesian approach to multicenter sparse data. *Communications in Statistics. Simulation and Computation*, *37*(4), 687-696. http://doi.org/10.1080/03610910701884062.

Tanizaki, H. (1997). Power comparison of non-parametric tests: small sample properties from Monte Carlo experiments. *Journal of Applied Statistics*, *24*(5), 603-632. http://doi.org/10.1080/02664769723576.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics*, *11*(2), 147-162. http://doi.org/10.1214/aoms/1177731909.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80-83. http://doi.org/10.2307/3001968.

Yang, G., Jiang, W., Yang, Q., & Yu, W. (2015). PBOOST: a GPU-based tool for parallel permutation tests in genome-wide association studies. *Bioinformatics*, *31*(9), 1460-1462. http://doi.org/10.1093/bioinformatics/btu840. PMid:25535244.

Zelterman, D. (1987). Goodness-of-Fit tests for large sparse multinomial distributions. *Journal of the American Statistical Association*, *82*(398), 624-629. http://doi.org/10.1080/01621459.1987.10478475.

Author Contributions

Jose Luiz Contador: Conceptualization, Methodology, Writing – original draft, Writing – review & editing
Edson Luiz França Senne: Software, Validation, Writing – review & editing
Jose Celso Contador: Validation, Writing – review & editing