CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open access journal indexed by

Volume 12 Number 1
April 2021
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Chilean Journal of Statistics

Volume 12, Number 1

April 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

Chilean Journal of Statistics
Vol. 12, No. 1, April 2021, 23–35

Non-parametric Inference
Research Paper

Exact tables for the Friedman rank test: Case with ties

Carlos López-Vázquez[1,*], Andrómaca Tasistro[2], and Esther Hochsztain[3]

[1]LatinGEO Lab IGM+ORT, Universidad ORT, Uruguay
[2]Permanent Education Unit, Faculty of Economic Sciences and Management, Universidad de la República, Uruguay
[3]Faculty of Economic Sciences and Management, Universidad de la República, Uruguay

(Received: 28 July 2020 · Accepted in final form: 01 April 2021)

Abstract

Exact tables for the case without ties of the Friedman statistic test proposed have been available since its inception. A modified statistic suitable for the case with ties has been derived 30 years later, and it appears in a text book nearly after 40 years. However, exact tables for the case of ties were never offered. Here we present for the first time a reduced set of exact tables for such a case, thus filling a gap. If a problem allows ties, the proper exact tables should be used thus disregarding other workarounds commonly suggested in the literature. The availability of exact tables for the case of ties is relevant for applied research because an hypothesis test decision when ties occur may be different if tables for the case without ties are used instead. We illustrate the effect of using the correct tables with both an example and a real data case study in the context of geoportals navigation analysis.

Keywords: Friedman test · Exact distribution · Non-parametric methods · R software.

Mathematics Subject Classification: Primary 62G10 · Secondary 62F05.

1. Introduction

The problem of analyzing the rankings resulting from a wine contest with k wines and N judges has been addressed by Friedman (1937). The null hypothesis is that there is no difference between the wines. In other situations, the wines might be medical treatments and the judges are patients. Original data might be of type ordinal, or it might be of continuous type (interval and ratio, as defined by Stevens, 1946). In that case, when ranking, he circumvents the normality requirements of other parametric tests like analysis of variance.

Tied ranks might appear with ordinal data, but also with continuous one. For example, if the values arise from a measurement device with finite accuracy, the uncertainty in their readings leads naturally to possible ties. In his seminal paper Friedman only considered the case without ties. Under such assumption he offered two asymptotic estimates valid for: a) large N irrespective of k and b) small N and moderate and large k. The asymptotic

*Corresponding author. Email: carloslopez@uni.ort.edu.uy

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

expressions were inaccurate for the case with both k and N small so he offered some exact tables. The set of $(k,\ N)$ pairs covered was modest, mostly due to the limited computing facilities of the time. The availability of exact tables and asymptotic approximations for ranking problems is not unusual; see Sen et al. (2011) for another example.

Exact tables for the case of ties were never published. In this proposal we offer for the first time some exact tables for the case of both small k and N. Exact tables for the Friedman rank test in the case of ties are relevant to applied research. In particular, we will show that the conclusion of an hypothesis test when ties occur may differ if tables for the case without ties are used instead of the ones proposed in this paper.

The paper is organized as follows. Section 2 describes the generalized formulae valid with or without ties as presented by various authors. On Section 3 we will comment about how the problem with ties is dealt with in textbooks, tutorials and reference guide. On principle, ties might be up to couples, triplets, etc. or without restrictions, something to be discussed in Section 4, showing that the results might vary depending on how many identical elements are allowed in the ties. Afterwards, in Section 5 an illustrative numerical example taken from a book will be presented followed by a case study in Section 6 showing the effect of misusing the asymptotic estimate and/or the wrong tables. The computational procedure will be commented in Section 7, and the new tables are presented in Section 8. Finally, some conclusions are sketched in Section 9.

2. Related works

2.1 The case without ties

Friedman (1937) proposed a rank test to avoid normality assumptions in analysis of variance. The method of ranks involves two steps, creating a two-way table: 1) rank data in each row 2) test if the all columns come from the same universe. The null hypothesis is that there are no differences between the columns. It can be proved that the statistic given by

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} \left(\sum_{i=1}^{N} r_{ij} \right)^2 - 3N(k+1)),$$

is asymptotically χ_{k-1}^2 distributed for large N, and to a normal distribution with mean $k-1$ and standard deviation $\sqrt{2(k-1)(N-1)/N}$ for small N and moderate to large k. The asymptotic estimate must be used if the (k, N) pair of the problem under consideration is not covered by the available tables. The jump between table values and asymptotic estimate might be large, so the set of exact tables for given (k, N) pairs steadily grew with time. Friedman (1937) considered only the cases of $k = 3$ for N up to 9, and $k = 4$ for N up to 4. Kendall and Babington-Smith (1939) extended the tables for $k = 3$ up to $N = 10$; $k = 4$ up to $N = 6$ and analyzed the case $k = 5$, $N = 3$. Owen (1962) published exact tables for $k = 3$ and N up to 15; $k = 4$ and N up to 8 without disclosing the computation procedure. Hollander and Wolfe (1999) provided tables for $k = 5$ and N up to 5. Odeh (1977) extended the tables considering the cases $k = 5$ for N up to 8, and $k = 6$ for N up to 6. Martin et al. (1993) extended the case $k = 4$ up to $N = 15$. A significant contribution was provided by van de Wiel (2000), who extended previous work considering the case $k = 3$ with N up to 25, $k = 4$ for N up to 20 and for $k = 5$ he offered tables for N up to 12. More recently López-Vázquez and Hochsztain (2019) drastically expanded the set of tables using a code in the `R` software implemented by Schneider et al. (2016). There, exact tables up to $N = 204$, 41, 13, 7, 4 and 2 for $k = 3$, 4, 5, 6, 7 and 8 were performed based on an algorithm proposed by van de Wiel et al. (1999). All of these efforts were for the case without ties.

2.2 The case with ties

In problems with discrete random variables ties are likely to appear, a fact acknowledged in Friedman (1937). It is difficult to explain why a correction for ties only was mentioned for the first time in Marascuilo and McSweeney (1967), and appeared significantly later in a textbook in Conover (1980). Other equivalent expressions appeared even later, like the one by Siegel and Castellan (1988) or the one proposed by Corder and Foreman (2009). Apparently they were derived independently, even though they produced exactly the same value. Unlike the simple case without ties, the rank for the case with ties has more than one alternative. Most of the literature used the mid-rank method, which assures that the sum of ranks for each judge is constant. According to Sprent and Smeeton (2007) the generalized statistic (now valid either with or without ties) still has the same asymptotic estimate as the original one proposed by Friedman (1937). However, despite they acknowledge that for low k and N the asymptotic estimate is not accurate, no exact tables were provided.

The correction for ties proposed by Marascuilo and McSweeney (1967) is stated as

$$\chi_r^2 = \frac{\frac{12}{k(k+1)} \sum_{j=1}^{k} \frac{R_j^2}{N} - 3N(k+1)}{1 - \frac{\sum_{s=1}^{d} (t_s^3 - t_s)}{N(k^3-k)}}.$$

The numerator is the familiar statistic for the case without ties. For the correction term, d is the number of set of ties and t_i is the number of tied scores in the i-th set of ties. If there are no ties the denominator is 1. The expression was not widely cited in the literature. Conover (1980) proposed an expression defined as

$$\chi_r^2 = \frac{(k-1) \sum_{j=1}^{k} \left[R_j - N(k+1)/2\right]^2}{\sum_{i=1}^{N} \sum_{j=1}^{k} {r_{ij}}^2 - Nk(k+1)^2/4},$$

where R_j is the sum of the ranks r_{ij} for treatment j. Apparently, a number of alternative formulations for the same statistic were derived independently. Siegel and Castellan (1988) proposed a slightly more complicated expression given by

$$\chi_r^2 = \frac{12 \sum_{j=1}^{k} R_j^2 - 3N^2k(k+1)^2}{Nk(k+1) + \frac{\left(Nk - \sum_{i=1}^{N} \sum_{j=1}^{g_i} t_{i,j}^3\right)}{k-1}},$$

where g_i is the number of sets of tied ranks in the i-th group and $t_{i,j}$ is the size of the j-th set of tied ranks in the i-th group.

Using the same definition for g_i and $t_{i,j}$, Hollander and Wolfe (1999) proposed a different expression, given by

$$\chi_r^2 = \frac{12 \sum_{j=1}^{k} R_j^2 - 3N^2k(k+1)^2}{Nk(k+1) - \frac{\sum_{i=1}^{N} \left\{ \left(\sum_{j=1}^{g_i} t_{i,j}^3 \right) - k \right\}}{k-1}}.$$

Gibbons and Chakraborti (2010) suggested another expression to be written as

$$S = \sum_{i=1}^{N}\left[\sum_{j=1}^{k}\left(r_{ij} - \frac{(N+1)}{2}\right)\right]^2; \quad \chi_r^2 = \frac{12(N-1)S}{Nk(N^2-1) - \sum\limits_{i=1}^{N}\sum\limits_{j=1}^{g_i}(t_i^3 - t_i)},$$

Buskirk et al. (2013) included other notation established as

$$\chi_r^2 = \frac{12\sum\limits_{j=1}^{k} R_j^2 - 3N^2k(k+1)^2}{Nk(k+1) + \frac{\sum_{i=1}^{N} t_i^3 - t_i}{k-1}},$$

with t_i being the number of observations involved in a tie for the i-th case.

Boos and Stefanski (2013) proposed a compact expression, now without the need to count the number of ties explicitly, given by

$$\chi_r^2 = \frac{(k-1)N^2\sum\limits_{j=1}^{k}\left[\frac{1}{N}\sum\limits_{i=1}^{N} r_{ij} - \frac{(k+1)}{2}\right]^2}{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{k} r_{ij}^2 - Nk(k+1)^2/4}.$$

In our computations, we use the expression from Corder and Foreman (2009), which is equivalent to the earlier ones, stated by

$$\chi_r^2 = \frac{N(k-1)\left[\sum\limits_{j=1}^{k}\frac{R_j^2}{N} - C_F\right]}{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{k} r_{ij}^2 - C_F},$$

where

$$C_F = \frac{Nk(k+1)^2}{4}.$$

3. Recommended strategies for the problem with ties

Without going as deep as Richardson (2019), who compared many aspects of non-parametric statistics textbooks, it is fit to mention how they treated the case of the Friedman test with ties. Our search included some of the books considered by Richardson (2019), all of them intended for a statistical audience, but also some others designed with other communities in mind. An example for the food industry might be Granato et al. (2014) and we have included some others in the list. The frontline of science is usually found at papers, not books. However, papers are typically known only to a very small community. Also, the textbooks offer guidelines to a variety of users, not necessarily experts in the field. Thus, it is important to assess how the case of the Friedman test with ties is considered in the literature intended to reach a large audience.

According to the literature, the alternatives at hand for a problem with ties are:

- Use the generalized statistic, and compare it against the asymptotic estimate based upon the χ^2 approximation thus ignoring the need of exact tables (Plichta and Garzon, 2009; Alvo and Yu, 2010; Sheskin, 2011; Vidakovic, 1999; Hettmansperger and McKean, 2011; Buskirk et al., 2013; Boos and Stefanski, 2013; Granato et al., 2014).
- Use the generalized statistic, and acknowledge that the χ^2 approximation will not be accurate for low k and N. Use as a surrogate the exact table without ties (van Belle et al., 2004; Zar, 2010; Gibbons and Chakraborti, 2010; Linebach et al., 2014; Chechile, 2020).
- Same as before, but noticing that they lack an exact table for the problem with ties (Sprent and Smeeton, 2007; Hollander et al., 2014).
- Ignore the effects of ties, and use both the traditional Friedman statistic as well as its associated exact tables (Greene and D'Oliveira, 2005; Daniel and Cross, 2013; Corder and Foreman, 2009, 2014).
- Break the ties, assigning random ranks through a Monte Carlo experiment, and then use the standard Friedman statistic (Rayner et al., 2005).
- Assume that the problem with ties can be handled just by using midranks (Lehman, 1975; Canavos, 1988; Derrac et al., 2011; Liu et al., 2017).

There is no good reason to keep using the traditional statistic as proposed by Friedman, because the generalized one considers both situations. However, neglecting the fact that the asymptotic estimate is only valid for mid to large k and N (López-Vázquez and Hochsztain, 2019), or that the exact tables are not valid for cases with ties might have a devastating effect on the conclusions. We will illustrate it with some cases, but before let us analyze a somewhat subtle detail.

4. Types of ties allowed

In general, ties might involve $2, \cdots, k$ wines, denoted here as p-tuples. The case without ties is equivalent to set $p = 1$. If, for some reason the problem of interest just allows ties of pairs but not triplets, we should use $p = 2$. The general case "with ties" is equivalent to set $p = k$. We have yet to find practical examples where p is not equal to either 1 or k, but if they exist the distinction might be important because the exact tables are different. We illustrate it in Table 1, which corresponds to the case $k = 5$ and $N = 4$. The possibilities range from $p = 1$ (denoted as "no ties") to $p = k = 5$ (denoted as "with ties").

Table 1. Effect on the critical values when ties are restricted to p-tuples. Case of $k = 5$ and $N = 4$.

	10	5	2.5	1.0	0.5	0.1
with ties	7.471	8.675	9.699	10.873	11.671	13.105
4-tuples	7.474	8.675	9.699	10.886	11.676	13.111
3-tuples	7.478	8.675	9.699	10.889	11.684	13.143
2-tuples	7.474	8.684	9.707	10.880	11.730	13.143
no ties	7.600	8.800	9.800	11.200	12.000	13.200

This has consequences not considered before by the literature. If the problem under consideration allows tied ranks, the corresponding exact tables must be used even if in the data under analysis there are no cases with ties. Thus, the strategy of "breaking the ties" with any procedure, thus reducing the problem to the case without ties, is flawed. The differences will be evident only for low k and N, when exact tables are needed. Otherwise, since the asymptotic estimate is the same for the case with or without ties there will be no difference.

5. Effect of misusing the standard tables for the problem with ties

Because it appears in a textbook we will first consider an example from Corder and Foreman (2009). They presented a simple example which is summarized in Table 2. It reports the ranks of the response of seven employees ($N = 7$) under three alternatives ($k = 3$) to deal with their tardiness: a) do nothing (denoted as baseline), b) one month with a monetary incentive of \$10, and c) another month with double incentive. They want to determine if either of the payback deduction strategies modified employee tardiness.

Table 2. Rank of tardiness after considering three incentive initiatives (from Corder and Foreman, 2009).

	Rank of monthly tardiness		
Employee	Baseline	Month 1	Month 2
1	2	3	1
2	3	2	1
3	2	3	1
4	3	2	1
5	3	1.5	1.5
6	2.5	2.5	1
7	3	2	1

In order to compute the statistic, we first find R_i summing table entries by columns as

$$\begin{aligned} R_1 &= 2+3+2+3+3+2.5+3 = 18.5 \\ R_2 &= 3+2+3+2+1.5+2.5+2 = 16 \\ R_3 &= 1+1+1+1+1.5+1+1 = 7.5 \end{aligned}$$

The denominator holds the sum of squares of the table entries as well as the C_F term

$$\begin{aligned} \sum_{i=1}^{7}\sum_{j=1}^{3} r_{ij}^2 = {} & 2^2+3^2+1^2+3^2+2^2+1^2+2^2+3^2+1^2+3^2+2^2+1^2+3^2+1.5^2+ \\ & +1.5^2+2^2+2.5^2+1^2+3^2+2^2+1^2 = 97, \end{aligned}$$

$$C_F = \frac{7*3*(3+1)^2}{4} = 84.$$

Thus the χ_r^2 value computation is presented as

$$\chi_r^2 = 7*2\left[\frac{18.5^2+16^2+7.5^2}{7} - 84\right] \Big/ (97-84) = \frac{133}{13} = 10.23.$$

For $\alpha = 0.05$ (Corder and Foreman, 2009) stated that the critical value is 7.140. After a quick check it is possible to confirm that the critical value presented corresponds to the "without ties" problem (see for example the tables from Martin et al., 1993). If we choose not to use exact tables, the χ^2 asymptotic approximation provides a critical value of $\chi^2_{0.05,2} = 4.605$. The correct value for the "with ties" problem should have been 5.769, taken from our Table 5. In this case the null hypothesis would be rejected using either critical value, but the differences observed are not negligible. An interesting case arises for $\alpha = 0.005$; the "without ties" table offers 10.286 as the critical value. The critical value χ^2 is now $\chi^2_{0.005,2} = 10.597$. Both are larger than the χ_r^2 so according to this the null hypothesis should be rejected. However, using our Table 5 the exact critical value is 9.083, now lower than the statistic value $\chi^2_r = 10.23$. Thus, according to the correct table, we could not reject the null hypothesis.

6. Case Study

We present a case study in the context of geoportals navigation analysis. As stated by Jiang et al. (2020) and Bernabé-Poveda and González (2014) a geoportal is a web-based solution to provide open spatial data sharing and online geo-information management. The concept of geoportals has becomed key for spatial data and geoinformation accessing and sharing. We perform geoportal navigation analysis based on geoportal web server logs (click-stream data) following the guidelines given by (Markov and Larose, 2007; Bhavani el al., 2017; Bhuvaneswari and Muneeswaran, 2021).

As indicated by Srivastava et al. (2019), whenever a user requests a particular web resource on a website, an entry is recorded into a log file which is automatically stored and maintained by the web server. The log file is named web server log or clickstream. We preprocessed web server logs and computed three variables to be used in this case study: pages per session, session duration and average time per page.

The double-entry table presented in Table 3 shows the coefficient of variation (rounded to two digits) of pages per session (CVPPS), where rows represent four levels I to IV of session duration and columns represent three levels I to III of average time per page. Levels were defined by percentile groups. Rankings by row are presented in Table 4, and we can observe that one tie occurs for Session duration level II. Figure 1 shows the R output of the Friedman test.

We want to assess if there is some relationship between the CVPPS and the session duration levels.

Table 3. Coefficient of variation of pages per session by session duration and average time per page levels.

Session duration levels	Coefficient of variation Average time per page levels		
	I	II	III
I	0.12	0.08	0.19
II	0.11	0.21	0.21
III	0.08	0.19	0.26
IV	0.13	0.22	0.27

Table 4. Coefficient of variation of pages per session by session duration and average time per page levels.

Session duration levels	Ranking of coefficient of variation of pages per session Average time per page levels		
	I	II	III
I	2	1	3
II	1	2.5	2.5
III	1	2	3
IV	1	2	3

```
> rankeddata
      [,1] [,2] [,3]
[1,]     2  1.0  3.0
[2,]     1  2.5  2.5
[3,]     1  2.0  3.0
[4,]     1  2.0  3.0
> friedman.test(rankeddata)
          Friedman rank sum test
data: rankeddata
Friedman chi-squared = 5.7333, df = 2, p-value = 0.05689
```

Figure 1. Friedman test output using R 3.6.3.

Following the same steps as before, we compute intermediate values and the resulting statistic, considering correction by ties as

$$\begin{aligned} R_1 &= 2+1+1+1 = 5 \\ R_2 &= 1+2.5+2+2 = 7.5 \\ R_3 &= 3+2.5+3+3 = 11.5 \end{aligned}$$

$$\sum_{i=1}^{4}\sum_{j=1}^{3} r_{ij}^2 = 2^2+1^2+1^2+1^2+1^2+2.5^2+2^2+2^2+3^2+2.5^2+3^2+3^2 = 55.5,$$

$$C_F = \frac{4*3*(3+1)^2}{4} = 48.$$

Thus, the χ_r^2 value is

$$\chi_r^2 = 4*(3-1)\left[\frac{5^2+7.5^2+11.5^2}{4} - 48\right] \Big/ (55.5-48) = \frac{43}{7.5} = 5.733,$$

The null hypothesis is that there are no differences between the columns, that is, for any given level of session duration, the coefficient of variation of pages per session (CVPPS) is the same at all levels of average time per page.

Friedman χ_r^2 statistic value is 5.733. At 5% significance level, if we consider exact tables without ties as provided by López-Vázquez and Hochsztain (2019) or Martin et al. (1993) the critical value for k=3 N=4 is 6.500, and therefore the decision is not to reject the null hypothesis. The same holds true if we consider the $\chi^2_{0.05,2}$ value for $k = 3$ (7.815), as the statistic value (5.733) is again less than the critical value. Considering the R-output shown in Figure 1, as p-value 0.05689 is larger than 0.05 we should conclude that the decision is not to reject the null hypothesis. However, when there are ties, as we can see in Table 3 it is necessary to use the exact tables presented in this paper. As we can see in Table 5, the correct critical value in this case at the 5% significance level is 5.571, and as a consequence the decision is to reject the null hypothesis. Therefore, using either the wrong exact table or the chi-square approximation results in a different decision than using the correct exact table presented in this paper. We can acknowledge the practical importance of using the proper tables for the Friedman Rank-Test in the case of ties, leading to a correct decision in the hypothesis test. And thus concluding that the different values of CVPPS have an effect over the Session duration levels.

7. Computational procedure to produce the tables

The problem of computing the exact tables for this problem has been barely considered in the literature. To the best of our knowledge, only Hollander and Wolfe (1999) proposed a brute-force procedure to compute the exact conditional distribution of the Friedman statistic valid for each particular case. We used instead a two step procedure in order to compute the general, exact tables, when there are ties among the data values. Firstly, we built the set of possible cases considering ties (always allowing up to k-tuples), and as a second step we computed the statistic for all the valid combinations. Hence, this is thus a combinatorial problem. For very small k computing the first step posed no special requirements. The computation was carried using the `R` software version 4.0.0 in a personal computer. The computation time for the second step was manageable for very small k and N, but the runtime requirements quickly grow along k and N. For example, for the case of $k = 4$, $N = 7$ the computations using Octave 3.8.2 required over 90 days of wall time using up to 100 nodes in parallel. The valid combinations were arranged in sets and computed independently using a nearly embarrassingly parallel approach. It is worth mentioning that, due to the combinatorial nature of the problem, our approach is only capable of dealing with modest cases in reasonable time. The facility (described by Nesmachnow and Iturriaga, 2019) is a LINUX-based cluster equipped with Intel Xeon-Gold 6138 2.00GHz processors.

8. Resulting tables

Tables 5, 6, 7, 8, and 9 are offered for the case of $k = 3, 4, 5, 6$ and 7, respectively. Following the style used by Martin et al. (1993), last row of each table holds the corresponding $\chi^2_{\alpha,k-1}$ asymptotic estimate. It illustrates the jump with respect to the exact table values for finite N.

Table 5. Critical values of the statistic χ^2_r in the case of ties up to k-tuples for $k = 3$ and N up to 11.

	0.100	0.050	0.025	0.010	0.005	0.001
$N = 3$	4.667	5.000	5.600	5.636	6.000	
$N = 4$	4.667	5.571	6.000	6.857	7.429	7.600
$N = 5$	4.588	5.647	6.615	7.444	8.316	9.294
$N = 6$	4.571	5.727	6.778	7.913	8.667	10.174
$N = 7$	4.560	5.769	6.870	8.222	9.083	10.583
$N = 8$	4.526	5.793	6.909	8.296	9.250	11.143
$N = 9$	4.563	5.813	6.968	8.357	9.455	11.467
$N = 10$	4.563	5.846	7.000	8.471	9.548	11.730
$N = 11$	4.550	5.850	7.048	8.581	9.657	12.000
$\chi^2_{\alpha,2}$	4.605	5.991	7.378	9.210	10.597	13.816

Table 6. Critical values of the statistic χ^2_r in the case of ties up to k-tuples for $k = 4$ and N up to 6.

	0.100	0.050	0.025	0.010	0.005	0.001
$N = 2$	5.400	5.842	5.842	6.000		
$N = 3$	5.893	6.692	7.444	8.111	8.379	8.793
$N = 4$	6.081	7.184	8.100	9.079	9.750	10.784
$N = 5$	6.136	7.370	8.478	9.652	10.421	11.936
$N = 6$	6.161	7.446	8.660	10.019	10.964	12.765
$\chi^2_{\alpha,3}$	6.251	7.815	9.348	11.345	12.838	16.266

Table 7. Critical values of the statistic χ_r^2 in the case of ties up to k-tuples for $k = 5$ and N up to 4.

	0.100	0.050	0.025	0.010	0.005	0.001
$N = 2$	6.703	7.263	7.568	7.789	7.897	8.000
$N = 3$	7.241	8.267	9.091	9.964	10.400	11.154
$N = 4$	7.471	8.675	9.699	10.873	11.671	13.105
$\chi^2_{\alpha,4}$	7.779	9.488	11.143	13.277	14.860	18.467

Table 8. Critical values of the statistic χ_r^2 in the case of ties up to k-tuples for $k = 6$ and N up to 3.

	0.100	0.050	0.025	0.010	0.005	0.001
$N = 2$	8.065	8.710	9.118	9.485	9.688	9.918
$N = 3$	8.580	9.697	10.637	11.634	12.255	13.284
$\chi^2_{\alpha,5}$	9.236	11.070	12.833	15.086	16.750	20.515

Table 9. Critical values of the statistic χ_r^2 in the case of ties up to k-tuples for $k = 7$ and $N = 2$.

	0.100	0.050	0.025	0.010	0.005	0.001
$N = 2$	9.303	10.073	10.624	11.094	11.345	11.725
$\chi^2_{\alpha,6}$	10.645	12.592	14.449	16.812	18.548	22.458

9. Conclusions

The Friedman rank test for the case without ties has been used for decades, but only until recently the case with ties was considered. Despite a correction for the original formulae is available, and that the asymptotic approximations are the same, no exact tables for low k and N have been published. Here we present the first ones, and illustrate its importance by showing that even a simple case published in a book suffers badly for using the wrong table in the computations. In addition, we consider that the tables for the case without ties are only applicable for problems which cannot accept ties, and not merely because the data do not show ties. This questioned some strategies that propose to break the ties reducing the problem to one without ties. In addition, the type of ties allowed (only pairs, only triplets, etc.) have a noticeable effect on the final decision, at least for small k and N. To build the exact tables we used a naive approach, which is combinatorial and can only deal with very small k and N. Further expansion of the exact tables will require using different algorithms, in the line of those of van de Wiel (2000) or van de Wiel et al. (1999).

Future works include expanding the exact tables and develop an `R` package to calculate Friedman rank test p-value based in the exact tables for the case of ties as described in this paper.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The experiments presented in this paper were carried out using ClusterUY (https://cluster.uy). We want to acknowledge Dr. Sergio Nesmachnow and Dr. Gabriel Usera for providing easy access to the computing facility. We also acknowledge Col. Norbertino Suárez and Lic. Sergio Marshall, both of the Military Geographic Institute, for providing access to the web server logs used in the case study.

References

Alvo, M. and Yu, P. L. H., 2014. Statistical Methods for Ranking Data. Springer, New York.

Bernabé-Poveda, M.A. and González, M.E., 2014. Sobre la necesaria usabilidad de los geoportales como puertas de entrada a las IDE. Revista Internacional de Ciencia y Tecnología de la Información Geográfica, 1(14), 1–5.

Bhavani, B and Sucharita, V. and Satyanarana, K.V.V., 2017. Review on techniques and applications involved in web usage mining. International Journal of Applied Engineering Research, 12, 15994–15998.

Bhuvaneswari, M.S. and Muneeswaran, K., 2021. User community detection from web server log using between user similarity metric. International Journal of Computational Intelligence Systems, 14, 266–281.

Boos, D.D. and Stefanski, L.A, 2013. Essential Statistical Inference: Theory and Methods. Springer, New York.

Buskirk, T. D.; Willoughby, L. M. and Tomazic, T.T., 2013. Nonparametric Statistical Techniques. Oxford Handbooks Online, Oxford, UK.

Canavos, G.C., 1988. Probability and Statistics: Applications and Methods. McGraw-Hill, Mexico.

Conover, W.J., 1980. Practical Nonparametric Statistics. Wiley, New York.

Corder, G.W. and Foreman, D.I., 2009. Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach. Wiley, NJ, US.

Corder, G.W. and Foreman, D.I., 2014. Nonparametric Statistics: A Step-by-Step Approach. Wiley, NJ, US.

Chechile, R.A., 2020. Bayesian Statistics for Experimental Scientists: A General Introduction Using Distribution-Free Methods. MIT Press, Massachusetts, US.

Daniel, W.W. and Cross, C.L., 2013. Biostatistics: A Foundation for Analysis in the Health Sciences. Wiley, New York.

Derrac, J., García, S., Molina, D., and Herrera, F., 2011. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm and Evolutionary Computation, 1, 3–18.

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, 675–701.

Gibbons, J. D. and Chakraborti, S., 2010. Nonparametric Statistical Inference. CRC Press, Boca Raton, FL, US.

Granato, D., Calado, V.M.A., and Jarvis, B., 2014. Observations on the use of statistical methods in food science and technology. Food Research International, 55, 137–149.

Greene, J. and D'Oliveira, M., 2005. Learning to Use Statistical Tests in Psychology. Open University Press.

Hettmansperger, T.P. and McKean, J.W., 2011. Robust Nonparametric Statistical Methods. CRC Press, Boca Raton, FL, US.

Hollander, M. and Wolfe, D.A, 1999. Nonparametric Statistical Methods. Wiley, New York.

Hollander, M., Wolfe, D.A. and Chicken, E., 2014. Nonparametric Statistical Methods. Wiley, New York.

Jiang, H., van Genderen, J., Mazzetti, P., Koo, H. and Chen, M., 2020. Current status and future directions of geoportals. International Journal of Digital Earth, 13, 1093–1114.

Kendall, M.G. and Babington-Smith, B., 1939. The problem of m rankings. Annals of Mathematical Statistics, 10, 275–287

Korosteleva, O., 2013. Nonparametric Methods in Statistics with SAS Applications. Chapman and Hall, New York.

Lehman, E.L. and D'Abrera, H.J.M., 1975. Nonparametrics. Statistical Methods Based on Ranks. Holden-Day, New York.

Linebach, J.A., Tesch, P.T., and Kovacsiss, L.M., 2014. Nonparametric Statistics for Applied Research. Springer, New York.

Liu, Z., Blash, E., and John, V., 2017. Statistical comparison of image fusion algorithms: Recommendations. Information Fusion, 36, 251–260.

López-Vázquez, C. and Hochsztain, E., 2019. Extended and updated tables for the Friedman rank test. Communications in Statistics: Theory and Methods, 48, 268–281.

Lovie, P., 2005. Coefficient of Variation. Encyclopedia of Statistics in Behavioral Science. Wiley, Online Library.

Marascuilo, L.A., and McSweeney, M., 1967. Nonparametric post-hoc comparisons for trend. Psychological Bulletin, 67, 401–412.

Markov, Z. and Larose, D.T., 2007. Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage. Wiley, New York.

Martin, L., Leblanc, R. and Toan, N., 1993. Tables for the Friedman rank test. Canadian Journal of Statistics, 21, 39–43.

Nesmachnow S. and Iturriaga S., 2019. Cluster-UY: collaborative scientific high performance computing in Uruguay. In: Torres M., Klapp J. (eds) Supercomputing. ISUM 2019. Communications in Computer and Information Science, vol 1151. Springer, Cham

Odeh, R.E., 1977. Extended tables of the distribution of Friedman's s-statistic in the two way layout. Communications in Statistics: Simulation and Computation, 6, 29–48.

Owen, D.B., 1962. Handbook of Statistical Tables. Addison-Wesley Publishing Company, Reading, Massachusetts, US.

Pélabon, C., Hilde, C., Einum, S. and Gamelon, M., 2020. 0n the use of the coefficient of variation to quantify and compare trait variation. Evolution Letters, 4, 180–188.

Plichta, S.B. and Garzon, L.S., 2009. Statistics for Nursing and Allied Health. Wolters Kluwer Health /Lippincott, Williams and Wilkins.

Rainer, J.C.W., Best, D.J., Brockhoff, P.B., and Rayner, G.D., 2005. Nonparametrics for Sensory Science: A More Informative Approach. Blackwell Publishing.

Richardson, A., 2019. A comparative review of nonparametric statistics textbooks. The American Statistician, 73, 360–366.

Schneider, G., Chicken, E., and Becvarik, R., 2016. NSM3 Package: Functions and Datasets to Accompany Hollander, Wolfe, and Chicken - Nonparametric Statistical Methods. R software v. 1.16

Sen, P.K., Salama, I.A., and Quade, D., 2011. Spearman's footrule: Asymptotics in applications. Chilean Journal of Statistics, 2(1), 3–20.

Sheskin, D.J., 2011. Handbook of Parametric and Nonparametric Statistical Procedures. Chapman-Hall, Boca Raton, FL, US.

Siegel, S. and Castellan, N.J.Jr., 1988. Nonparametric Statistics for the Behavioral Sciences. Mcgraw-Hill, New York, US.

Sprent, P. and Smeeton, N., 2007. Applied Nonparametric Statistical Methods. CRC, NY.

Srivastava, M., Srivastava; A.K. and Garg, R., 2019. Data preprocessing techniques in web usage mining: A literature review. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), pp. 466-476

Stevens, S.S., 1946. On the theory of scales of measurement. Science, 103, 677–680.

van Belle, G., Fisher, L.D., Heagerty, P.J., and Lumley, T., 2004. Biostatistics: A Methodology for the Health Sciences. Wiley, New Work, US.

van de Wiel, M.A., 2000. Exact distributions of distribution-free test statistics. Ph.D. Thesis, Eindhoven University of Technology, The Netherlands.

van de Wiel, M.A., Di Bucchianico, A. and van der Laan, P., 1999. Symbolic computation and exact distributions of nonparametric test statistics. Journal of the Royal Statistical Society D, 48, 4, 507–516.

Vidakovic, B., 2011. Statistics for Bioenginering Sciences with MATLAB and WinBUGS Support. Springer, New Work, US.

Zar, J.H., 2010. Biostatistical Analysis. Pearson, India.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright

Authors who publish their articles in the ChJS automatically transfer their copyright to the Chilean Statistical Society. This enables full copyright protection and wide dissemination of the articles and the journal in any format. The ChJS grants permission to use figures, tables and brief extracts from its collection of articles in scientific and educational works, in which case the source that provides these issues (Chilean Journal of Statistics) must be clearly acknowledged.