Scientific
Research
Publishing

Adjustments for Kurtosis and Continuity on the Prentice Test

Lily Gebhart[1], John Kolassa[2]

[1]Math Department, Occidental College, Los Angeles, USA
[2]Department of Statistics, University of New Jersey, Piscataway, USA
Email: kolassa@stat.rutgers.edu

Abstract

The test of Prentice [1] is a non-parametric statistical test for the two-way analysis of variance using ranks. The null distribution of this test typically is approximated using the Chi-square distribution. However, the exact null distribution deviates from the Chi-square approximation in certain cases commonly found in applications of the test, motivating adjustments to the distribution. This manuscript presents adjustments to this null distribution correcting for continuity, multivariate skewness, and multivariate kurtosis. The effects of alternative scoring methods as non-polynomial functions of rank sums are also presented as a broader application of the approximation.

Keywords

Friedman Test, Prentice Test, Edgeworth Approximation

1. Introduction

The Prentice test [1] is the nonparametric analog of a two-way ANOVA, widely used in survival analysis, agricultural studies, and more generally, in biostatistics. The test is particularly useful for analyzing data that do not necessarily follow a normal distribution since ranking the data removes dependence on the original distribution. This method can be applied to blocked data of several treatments with variable and potentially unbalanced replicates corresponding to each block and treatment combination. Several special cases exist, including the Kruskal-Wallis Test, the nonparametric analog of a one-way ANOVA with one block and variable replicates [2], and the Friedman test, the case of the Prentice test with one replicate per group-block combination [3]. The special cases of the Prentice test, as well as its nonparametric nature and adjustments for unbalanced replicates,

render the test flexible and applicable to analyzing a wide range of data. These features of the Prentice test are particularly significant considering that few if any real-world datasets requiring statistical analysis are normally distributed and balanced due to participant dropout and noisy data commonplace in many practical applications.

Despite being an important statistical test, computing the exact Prentice test statistic distribution for practical applications is highly computationally expensive. Tables with Prentice test statistic values for small examples exist, but in most practical applications, the Prentice test statistic is applied to larger examples. Furthermore, the use of tables has the potential to lead to inaccurate conclusions, as values achieved in applications rarely match the specific test statistic values included in tables, in which interpolation or rounding can result in erroneous conclusions especially considering the discontinuous nature of the Prentice distribution.

Several approximations via less computationally expensive test distributions have been developed, namely the Chi-square distribution and the Iman-Davenport approximation, but they fail to fully capture the behavior of the Prentice test distribution, especially near the tail of the distributions. Since most practical applications require test statistic values from the tail of the distribution, inaccurate approximations can lead to false conclusions which may result in devastating consequences.

The null distribution of the Prentice test and its special cases are commonly approximated by the Chi-square distribution. Other multinomial test statistics, most notably the generalized likelihood ratio statistic, are not considered in this manuscript [4].

Bounds on the Chi-square approximation to the Friedman test were produced for both central and non-central distributions and under the null and alternative hypotheses. The general bounds are of order $o\left(N^{-1/2}\right)$ and in the central case, bounds are improved to order $o\left(N^{-(k-1)/k}\right)$ for the Chi-square distribution with $k-1$ degrees of freedom [5]. More recent bounds on the Chi-square approximation to the Prentice test statistic have been produced using Stein's method, originally utilized for approximating the distance between the normal distribution and a probability distribution of choice, but which have also been applied to bounding approximations to the χ^2 distribution [6]. For k treatments and b blocks, the distance between the Prentice test statistic distribution and the Chi-square distribution with $k-1$ degrees of freedom is bounded by order b^{-1} [7]. Furthermore, the bound is dependent on k, approaching zero only if k/b also approaches zero [7].

Limitations to the approximation by the Chi-square distribution result from the continuity of the distribution and the assumption that the parameters in the multinomial distribution studied are independent and identically distributed [4]. The dependence of the Chi-square approximation on the number of blocks and treatments as well as the limitations of its i.i.d. assumption will be presented via

example in the sections to follow.

To date, several improvements have been made to the approximation of the Friedman and Kruskal Wallis test statistics. Of note is the *F* Statistic approximation, one of several approximations made by Iman and Davenport and referred to as the Iman-Davenport approximation throughout [8] [9]. While the Chi-square approximation frequently underestimates the critical region of the Friedman test statistic, the Iman-Davenport approximation frequently overestimates the critical region making it a useful comparison [8].

Here, we apply the adjustments to the Chi-square distribution presented by Yarnold to the Kruskal-Wallis, Friedman, and Prentice tests. The approximation applied results from the integration of an Edgeworth asymptotic expansion for $Pr(T \in B)$ where B is a Borel set and T the groupwise sums of k independent random vectors. When B is the ellipse corresponding to the critical region for the Prentice test, and the Edgeworth approximation is integrated, the resulting approximation consists of the adjustments to the Chi-square distribution function for continuity and kurtosis, respectively [10] [11]. When applied to the Kruskal-Wallis, Friedman, and Prentice test statistics, the adjustments introduced by Yarnold provide significant corrections to the Chi-square distribution function approximation for each test statistic distribution.

Notably, the corrections that the Yarnold approximation yields for continuity and multivariate kurtosis provide a more accurate representation of the tail probabilities of the Prentice test distribution than previous approximations. The adjustment for continuity provides a more accurate representation of the discontinuous behavior of the Prentice distributions than previous approximations, where i.i.d. assumptions result in continuous approximations. Furthermore, the adjustment for kurtosis in the Yarnold approximation more accurately reflects the distribution of probability in the tail versus the center of the Prentice distribution, resulting in better approximations to the tail of the distribution, which is especially useful for practical applications of the test. These improvements are also applicable to all subcases of the Prentice test, which enables more accurate data interpretation in the diverse research context of the Prentice test commonly used.

2. Methods

Let T be a random variable defined as a function of rank sums with a distribution of k degrees of freedom. Let κ_2, κ_3, and κ_4 denote the second, third, and fourth multivariate cumulants respectively. The cumulants are calculated from the computed central moments of the test statistics, and depend on the number of groups, replicates, and blocks in the design using the algebraic relationship between central moments and cumulants [12].

2.1. The Yarnold Approximation

The approximation by Yarnold is applied to improve approximations for the Kruskal-Wallis, Friedman, and Prentice tests. The second and third partial sums

of the Yarnold approximation were considered separately as approximation A and approximation B. Approximation A corrects for continuity and approximation B corrects for both continuity and kurtosis. Here, approximation A is valid to $O\left(\dfrac{1}{n}\right)$ and it is conjectured, but not proven, that approximation B is valid to $o\left(\dfrac{1}{n}\right)$ [10]. Approximations A and B are presented in Equations (1) and (2), respectively [10].

$$Pr(T < c) \approx \chi_k^c + \left(N(nc) - V(nc)\right)\frac{\mathrm{e}^{-c/2}}{(2\pi n)^{k/2}|\kappa_2|^{1/2}} \tag{1}$$

$$
\begin{aligned}
Pr(T < c) \approx{}& \chi_k^c + \left(N(nc) - V(nc)\right)\frac{\mathrm{e}^{-c/2}}{(2\pi n)^{k/2}|\kappa_2|^{1/2}} \\
&- \left[\frac{\delta_1}{n}\left(\sum_{t=0}^{2}(-1)^{2-t}\binom{2}{t}\chi_{k+2t}^c\right) + \frac{\delta_2}{n}\left(\sum_{t=0}^{3}(-1)^{3-t}\binom{3}{t}\chi_{k+2t}^c\right)\right]
\end{aligned}
\tag{2}
$$

While the original approximation applied techniques to means of independent replicates, we apply the approximation to summaries with standardized cumulants that have the same structure [10]. Hence, we take $n = 1$. Here,

$$\delta_1 = \frac{1}{8}\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}\sum_{i_4=1}^{k}\left(\kappa_2^{-1}\right)^{i_1 i_2}\left(\kappa_2^{-1}\right)^{i_3 i_4}\kappa_4^{i_1 i_2 i_3 i_4}$$

and

$$
\begin{aligned}
\delta_2 ={}& \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}\sum_{i_4=1}^{k}\sum_{i_5=1}^{k}\sum_{i_6=1}^{k}\frac{\left(\kappa_2^{-1}\right)^{i_1 i_2}\left(\kappa_2^{-1}\right)^{i_3 i_4}\left(\kappa_2^{-1}\right)^{i_5 i_6}}{8} \\
&+ \frac{\left(\kappa_2^{-1}\right)^{i_1 i_4}\left(\kappa_2^{-1}\right)^{i_2 i_5}\left(\kappa_2^{-1}\right)^{i_3 i_6}}{12}\kappa_3^{i_1 i_2 i_3}\kappa_3^{i_4 i_5 i_6}.
\end{aligned}
$$

In the equation above, $N(nc)$ refers to the number of points on the lattice in the probability ellipse and $V(nc)$ refers to the volume of the probability ellipse [10]. For the test statistic T, the probability ellipse is

$$T = (Y - \mu)^{\mathrm{T}}\Sigma^{-1}(Y - \mu)$$

with Y the group rank sums of the test statistic, excluding one group, μ the expectation of Y, and Σ the null variance-covariance matrix of Y. See **Figure 1** for an example.

This approximation was applied to the Prentice test and compared to that of the Chi-square distribution with k degrees of freedom and the Monte Carlo evaluation of the true distribution of the Prentice test statistic under the assumption of treatment homogeneity. Here, both balanced and unbalanced cases with variable group and block counts were considered. Approximations to the Kruskal-Wallis and Friedman test statistics occur as special cases of the Prentice test approximation. The approximation to the Kruskal-Wallis test statistic occurs in cases when one block is considered and the approximation to the Friedman test statistic when one replicate per group-block combination is considered.
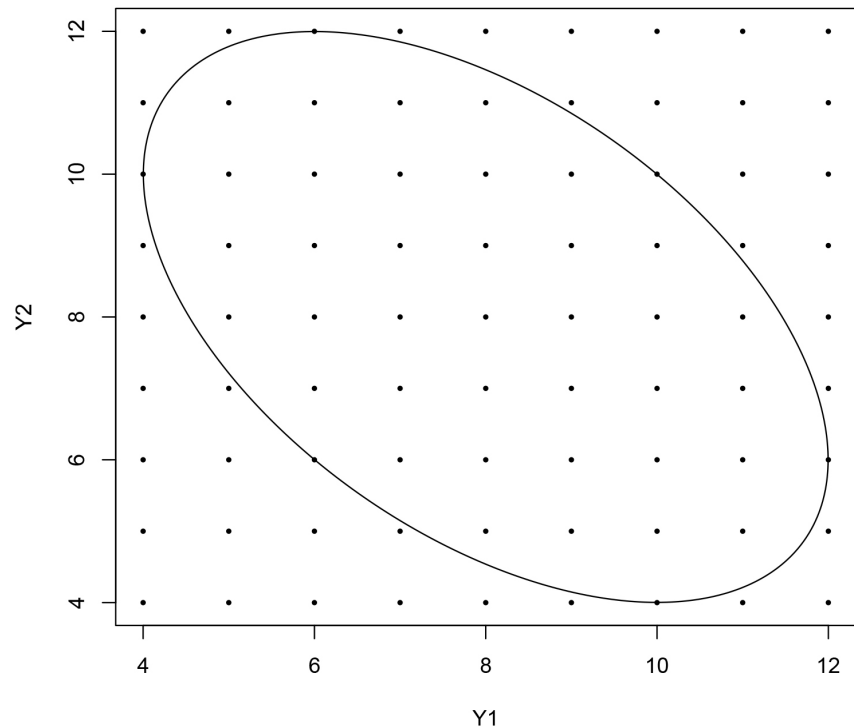
Figure 1. The ellipse for the Friedman test statistic in a case with three groups and four blocks with one replicate in each combination. The number of lattice points falling inside of the ellipse are summed in $N(nc)$ and the volume of the ellipse is expressed as $V(nc)$.

In the case of the Friedman and Kruskal-Wallis test statistics, another comparison is made with the Iman-Davenport approximation [8] [9].

To apply the Yarnold approximation with the homogeneity assumption to each test statistic, the average rank sums were computed for each specified number of groups, blocks, and replicates.

The Friedman, Prentice, and Kruskal-Wallis tests are generalizations of the Wilcoxon rank sum test. The Wilcoxon test is a member of larger family of general score statistics, formed by replacing the ranks by a monotonic transformation of ranks. Members of this family with scores other than the raw ranks can be chosen based on the expected distribution of errors. The particular choice of ranks as scores is optimal for Laplace errors [13].

Alternative scoring measures were also applied here, where the scores assigned to each item were non-polynomial functions of the ranks, namely logarithmic functions. The new scores were then summed by group, and the associated quadratic form was used as the test statistic. The application of the Yarnold approximation was otherwise unchanged.

The central moments and cumulants are calculated from the number of replicates in each block by treatment category, and are thus dependent on the case considered. Along with the degrees of freedom as $k - 1$, the second, third, and fourth cumulants and second central moment of each case enabled the Yarnold approximation to be tailored to each test.

2.2. Central Moments of Generalized Rank Statistics

Suppose that Y_{jk} is the rank sum for observations in group j and block $k \in \{1, \cdots, K\}$. Let $Y_{j.}$ be the sum of ranks in group i over all blocks; $Y_{j.} = \sum_{k=1}^{K} Y_{jk}$. Let Σ be the $J \times J$ matrix of variances and covariances for these rank sums; $\Sigma_{j\ell} = \text{Cov}\left[Y_{j.}, Y_{\ell.}\right]$. Let Λ represent the inverse of Σ with row and column J removed. Then the Prentice statistic is

$$\sum_{j=1}^{J-1} \sum_{\ell=1}^{J-1} \left(Y_{j.} - \text{E}\left[Y_{j.}\right]\right) \Lambda^{j\ell} \left(Y_{\ell.} - \text{E}\left[Y_{\ell.}\right]\right).$$

Let I_i^a be an indicator of whether the subject ranked i falls into group a. Consider the test statistic for group a, $X^a = \sum_{i=1}^{n} r_i I_i^a$, for scores r_j. The standard Wilcoxon rank sum statistic is given by $r_i = i$. Its centered version is given by $r_i = i - (n+1)/2$. Let $\sum^*$ represent summation over all sets of subscripts on ranks, omitting any with repeated values; then, for example, for any integers p and q, $\sum^* ij r_i^p r_j^q = \sum_{i=1}^{n} \sum_{j \neq i} r_i^p r_j^q$.

Second powers of the test statistic are given by $X^a X^b = \sum_{i=1}^{n} r_i I_i^a \sum_{j=1}^{n} r_j I_j^b$. Separating into sums without repeated indices,

$$X^a X^b = \sum_{i=1}^{n} r_i I_i^a \left(r_i I_i^b + \sum_{j \neq i}^{n} r_j I_j^b \right) = \sum^* r_i^2 I_i^a I_i^b + \sum^* r_i r_j I_i^a I_j^b.$$

The expectation of the sum is the sum of expectations, and so

$$\text{E}\left[X^a X^b\right] = \sum^* r_i^2 \text{E}\left[I_i^a I_i^b\right] + \sum^* r_i r_j \text{E}\left[I_i^a I_j^b\right].$$

Let μ with ordered subscripts and superscripts represent the expectation of the product of indicators; that is, for example, $\mu_{ij}^{ab} = \text{E}\left[I_i^a I_i^b\right]$. Then

$$\text{E}\left[X^a X^b\right] = \sum^* r_i^2 \mu_{ii}^{ab} + \sum^* r_i r_j \mu_{ij}^{ab}.$$

Because under the hypothesis of homogeneity, μ_{ij}^{ab} does not depend on the values of i and j so long as one keeps track of which of these are distinct, then $\text{E}\left[X^a X^b\right] = S_1^2 \mu_{11}^{ab} + S_2^2 \mu_{12}^{ab}$, for

$$S_1^2 = \sum^* r_i^2; \ S_2^2 = \sum^* r_i r_j.$$

When $r_i = i$,

$$S_1^2 = \frac{n(n+1)(2n+1)}{6}; \ S_2^2 = \frac{n(n+1)(3n^2 - n - 2)}{12}.$$

When $r_i = i - (n+1)/2$ then

$$S_1^2 = \frac{n(n^2 - 1)}{12}; \ S_2^2 = \frac{n(1 - n^2)}{12}.$$

Table 1 contains expectations of these indicators, depending on which group indicators are equal. A pattern with adjacent indicators indicates equality, and with bars between them inequality. The first row in this table represents the case in which $a = b$, and the second represents the case in which $a \neq b$.

Third powers of the test statistic are given by

Table 1. Expectations of products of two indicators.

Pattern	μ_{11}^{ab}	μ_{12}^{ab}
ab	$\dfrac{n_a}{n}$	$\dfrac{n_a(n_a-1)}{n(n-1)}$
$a\,\vert\,b$	0	$\dfrac{n_a n_b}{n(n-1)}$

$$X^a X^b X^c = \sum_{i=1}^{n} r_i I_i^a \sum_{j=1}^{n} r_j I_j^b \sum_{k=1}^{n} r_k I_k^c.$$

Separating into sums without repeated indices,

$$X^a X^b X^c = \sum{}^{*} r_i^3 I_i^a I_i^b I_i^c + \sum{}^{*} r_i^2 r_k I_i^a I_i^b I_k^c + \sum{}^{*} r_i^2 r_j I_i^a I_j^b I_i^c$$
$$+ \sum{}^{*} r_i r_j^2 I_i^a I_j^b I_j^c + \sum{}^{*} r_i r_j r_k I_i^a I_j^b I_k^c.$$

Then $\mathrm{E}\left[X^a X^b X^c\right] = S_1^3 \mu_{111}^{abc} + S_2^3 \mu_{112}^{abc} + S_2^3 \mu_{121}^{abc} + S_2^3 \mu_{211}^{abc} + S_3^3 \mu_{123}^{abc}$, for

$$S_1^3 = \sum{}^{*} r_i^3;\ \ S_2^3 = \sum{}^{*} r_i r_j^2;\ \ S_3^3 = \sum{}^{*} r_i r_j r_k.$$

When $r_i = i$,

$$S_1^3 = \frac{n^2(n+1)^2}{4};\ \ S_2^3 = \frac{n^2(n+1)(n^2-1)}{6};$$
$$S_3^3 = \frac{n^2(n+1)(n^3-2n^2-n+2)}{8}.$$

When $r_i = i - (n+1)/2$ then $S_m^3 = 0$ for $m = 1, 2, 3$.

Table 2 contains expectations of these indicators, depending on which group indicators are equal. A pattern with adjacent indicators indicates equality, and with bars between them inequality; note $a\,\vert\,bc$ represents $a \neq b = c$. The [3] in the heading to the column with $\mathrm{E}\left[I_1^a I_1^b I_2^c\right]$ represents the fact that a, b, and c can be matched with subjects 1 and 2 in three distinct ways; the column entries represent the sum of the three rearrangements. The first entry in this column has the multiplier 3, because all arrangements lead to the identical expectation when all groups are the same. The second entry lacks this multiplier, since it represents the case with two distinct groups; only the arrangement placing both with subject 1 into the same group represents a positive probability. The third entry is zero, since that entry represents the case with three distinct groups, and this cannot happen if subject 1 is assigned both to groups a and b.

Fourth powers of the test statistic are given by

$$X^a X^b X^c X^d = \sum_{i=1}^{n} r_i I_i^a \sum_{j=1}^{n} r_j I_j^b \sum_{k=1}^{n} r_k I_k^c \sum_{m=1}^{n} r_m I_m^d.$$

Separating into sums without repeated indices,

Table 2. Expectations of products of three indicators.

Pattern	μ_{111}^{abc}	μ_{112}^{abc} [3]	μ_{123}^{abc}
abc	$\dfrac{n_a}{n}$	$\dfrac{3n_a(n_a-1)}{n(n-1)}$	$\dfrac{n_a(n_a-1)(n_a-2)}{n(n-1)(n-2)}$
$a\,|\,bc$	0	$\dfrac{n_a n_b}{n(n-1)}$	$\dfrac{n_a n_b(n_b-1)}{n(n-1)(n-2)}$
$a\,|\,b\,|\,c$	0	0	$\dfrac{n_a n_b n_c}{n(n-1)(n-2)}$

$$
\begin{aligned}
X^a X^b X^c X^d = & \sum\nolimits^* r_i^4 I_i^a I_i^b I_i^c I_i^d + \sum\nolimits^* r_i^3 r_m I_i^a I_i^b I_i^c I_m^d + \sum\nolimits^* r_i^3 r_k I_i^a I_i^b I_k^c I_i^d \\
& + \sum\nolimits^* r_i^2 r_k^2 I_i^a I_i^b I_k^c I_k^d + \sum\nolimits^* r_i^2 r_k r_m I_i^a I_i^b I_k^c I_m^d + \sum\nolimits^* r_i^3 r_j I_i^a I_j^b I_i^c I_i^d \\
& + \sum\nolimits^* r_i^2 r_j^2 I_i^a I_j^b I_i^c I_j^d + \sum\nolimits^* r_i^2 r_j r_m I_i^a I_j^b I_i^c I_m^d + \sum\nolimits^* r_i^2 r_j^2 I_i^a I_j^b I_j^c I_i^d \\
& + \sum\nolimits^* r_i r_j^3 I_i^a I_j^b I_j^c I_j^d + \sum\nolimits^* r_i^2 r_j r_m I_i^a I_j^b I_j^c I_m^d + \sum\nolimits^* r_i^2 r_j r_k I_i^a I_j^b I_k^c I_i^d \\
& + \sum\nolimits^* r_i r_j^2 r_k I_i^a I_j^b I_k^c I_j^d + \sum\nolimits^* r_i r_j r_k^2 I_i^a I_j^b I_k^c I_k^d + \sum\nolimits^* r_i r_j r_k r_m I_i^a I_j^b I_k^c I_m^d .
\end{aligned}
$$

Taking expectations,

$$
\begin{aligned}
\mathrm{E}\left[X^a X^b X^c X^d\right] = & \sum\nolimits^* r_i^4 \mu_{1111}^{abcd} + \sum\nolimits^* r_i^3 r_j \mu_{1112}^{abcd} + \sum\nolimits^* r_i^3 r_j \mu_{1121}^{abcd} + \sum\nolimits^* r_i^2 r_j^2 \mu_{1122}^{abcd} \\
& + \sum\nolimits^* r_i^2 r_j r_k \mu_{1123}^{abcd} + \sum\nolimits^* r_i^3 r_j \mu_{1211}^{abcd} + \sum\nolimits^* r_i^2 r_j^2 \mu_{1211}^{abcd} + \sum\nolimits^* r_i^2 r_j r_k \mu_{1213}^{abcd} \\
& + \sum\nolimits^* r_i^2 r_j^2 \mu_{1221}^{abcd} + \sum\nolimits^* r_i r_j^3 \mu_{1222}^{abcd} + \sum\nolimits^* r_i^2 r_j r_m \mu_{1223}^{abcd} + \sum\nolimits^* r_i^2 r_j r_k \mu_{1231}^{abcd} \\
& + \sum\nolimits^* r_i r_j^2 r_k \mu_{1232}^{abcd} + \sum\nolimits^* r_i r_j r_k^2 \mu_{1233}^{abcd} + \sum\nolimits^* r_i r_j r_k r_m \mu_{1234}^{abcd} .
\end{aligned}
$$

For the centered scores $r_i = i - (n+1)/2$,

$$
S_1^4 = \sum\nolimits^* r_i^4 = \frac{n(3n^4 - 10n^2 + 7)}{240}
$$

$$
S_2^4 = \sum\nolimits^* r_i^3 r_j = \frac{(-3n^5 + 10n^3 - 7n)}{240}
$$

$$
S_3^4 = \sum\nolimits^* r_i^2 r_j^2 = \frac{n(5n^5 - 9n^4 - 10n^3 + 30n^2 + 5n - 21)}{720}
$$

$$
S_4^4 = \sum\nolimits^* r_i^2 r_j r_k = \frac{n(-5n^5 + 18n^4 + 10n^3 - 60n^2 - 5n + 42)}{720}
$$

$$
S_5^4 = \sum\nolimits^* r_i r_j r_k r_m = \frac{n(5n^5 - 18n^4 - 10n^3 + 60n^2 + 5n - 42)}{240} .
$$

Then

$$
\mathrm{E}\left[X^a X^b X^c X^d\right] = S_1^4 \mu_{1111}^{abcd} + S_2^4 \mu_{1112}^{abcd}[4] + S_3^4 \mu_{1122}^{abcd}[3] + S_4^4 \mu_{1123}^{abcd}[6] + S_5^4 \mu_{1234}^{abcd},
$$

Table 3 contains expectations of these indicators, depending on which group

Table 3. Expectations of products of four indicators.

Pattern	μ_{1111}^{abcd}	μ_{1112}^{abcd} [4]	μ_{1122}^{abcd} [3]	μ_{1123}^{abcd} [6]	μ_{1234}^{abcd}
$abcd$	$\dfrac{n_a}{n}$	$\dfrac{4n_a(n_a-1)}{n(n-1)}$	$\dfrac{3n_a(n_a-1)}{n(n-1)}$	$\dfrac{6n_a(n_a-1)(n_a-2)}{n(n-1)(n-2)}$	$\dfrac{n_a(n_a-1)(n_a-2)(n_a-3)}{n(n-1)(n-2)(n-3)}$
$a\,|\,bcd$	0	$\dfrac{n_a n_b}{n(n-1)}$	0	$\dfrac{3n_a n_b(n_b-1)}{n(n-1)(n-2)}$	$\dfrac{n_a n_b(n_b-1)(n_b-2)}{n(n-1)(n-2)(n-3)}$
$ab\,|\,cd$	0	0	$\dfrac{n_a n_b}{n(n-1)}$	$\dfrac{n_a n_b(n_a+n_b-2)}{n(n-1)(n-2)}$	$\dfrac{n_a(n_a-1)n_b(n_b-1)}{n(n-1)(n-2)(n-3)}$
$a\,|\,b\,|\,cd$	0	0	0	$\dfrac{n_a n_b n_c}{n(n-1)(n-2)}$	$\dfrac{n_a n_b n_c(n_c-1)}{n(n-1)(n-2)(n-3)}$
$a\,|\,b\,|\,c\,|\,d$	0	0	0	0	$\dfrac{n_a n_b n_c n_d}{n(n-1)(n-2)(n-3)}$

indicators are equal. A pattern with adjacent indicators indicates equality, and with bars between them inequality; note $a\,|\,b\,|\,cd$ represents $a \neq b \neq c = d$ and $a \neq c$.

3. Results

This section presents an illustrative example to demonstrate the improvements of our approximation on previous approximations and several cases to demonstrate the general applicability of our approximation.

3.1. Illustrative Example

Consider the effectiveness of advertising for a marketing firm via direct mail, newspaper, and magazine for twelve companies over the course of a year. In this example, each of the clients receives each advertising method over the course of a year and the Friedman test is run to discern the effects of the median response rate for each advertising method [14].

In this smaller example, the greater applicability of our approximation is better demonstrated. In these results in Table 4, our approximation results yields a conservative estimate of the critical value of the Prentice Test statistic, which we approximated via Monte Carlo simulation. However, the Chi-Square and Iman-Davenport approximations yield liberal estimates that are much further off from the accepted critical value.

3.2. General Cases

To demonstrate the applicability of our approximation, several cases are presented varying numbers of blocks and groups. In each case presented, plots comparing the distribution of the test statistic in comparison to other approximations and the error of the approximations relative to the Prentice test statistic will be presented from the 50th to the 99th quantile of the distribution of the Chi-square test statistic with $k-1$ degrees of freedom.

Table 4. Approximation results of the marketing firm example.

Approximation	Critical Value
Prentice Test (Monte Carlo)	0.94196
Chi-Square	0.95
Iman-Davenport	0.957544
Yarnold A	0.9391942
Yarnold B	0.9386689

The mean (Mean RE) and standard deviation (SD RE) of the error of each approximation relative to the Prentice test will also be presented with each example for comparison purposes.

Note that the scale for the relative error plot changes depending on the range of relative error observed in each case. Figure 2 displays a case with relatively low counts of groups, blocks, and replicates for comparison purposes.

Even in this small example, the Yarnold A (Mean RE 0.258, SD RE 0.2973824) and Yarnold B (Mean RE 0.239, SD RE 0.2577215) approximations yield a general improvement over the Chi-Square (Mean RE 0.312, SD RE 0.312 and Iman-Davenport (Mean RE 0.321, SD 0.255) approximations.

As will be displayed by the mean and standard deviation of the relative error of each approximation, generally, both approximations A and B improve as the counts of groups, blocks, and replicates increase, but becomes less differentiated from the Chi-square distribution. The approximation improves most markedly as the number of blocks increase.

Decreasing the number of blocks to 1, as shown in Figure 3, greatly reduces the accuracy of all approximations other than the Iman-Davenport approximation

(Mean RE 0.191, SD RE 0.174) specific to the Kruskal Wallis test [9]. Approximations A

(Mean RE 1.448, SD RE 2.811) and B

(Mean RE 1.461, SD RE 2.819) only have marginally lower relative error than the Chi-square distribution

(Mean RE 1.472, SD RE 2.825) However, the difference in relative error improves with larger sample sizes, as shown in Figure 4, where the replicates are increased from 3 to 10 in each group-block combination. In this case, all approximations are highly accurate with a small disparity between the Iman Davenport (Mean RE 0.048, SD RE 0.049) and Yarnold B (Mean RE 0.1088358, SD RE 0.108) approximations and the Chi-Square (Mean RE 0.115, SD RE 0.128) and Yarnold A approximations (Mean RE 0.115, SD RE 0.128).

Figure 5 displays the improvement of the approximation at high numbers of blocks, holding the replicate and group counts at relatively low values.

In this case, approximations A (Mean RE 0.0232, SD RE 0.02109596) and B (Mean RE 0.0202, SD RE 0.016) display marked improvements to Chi-Square
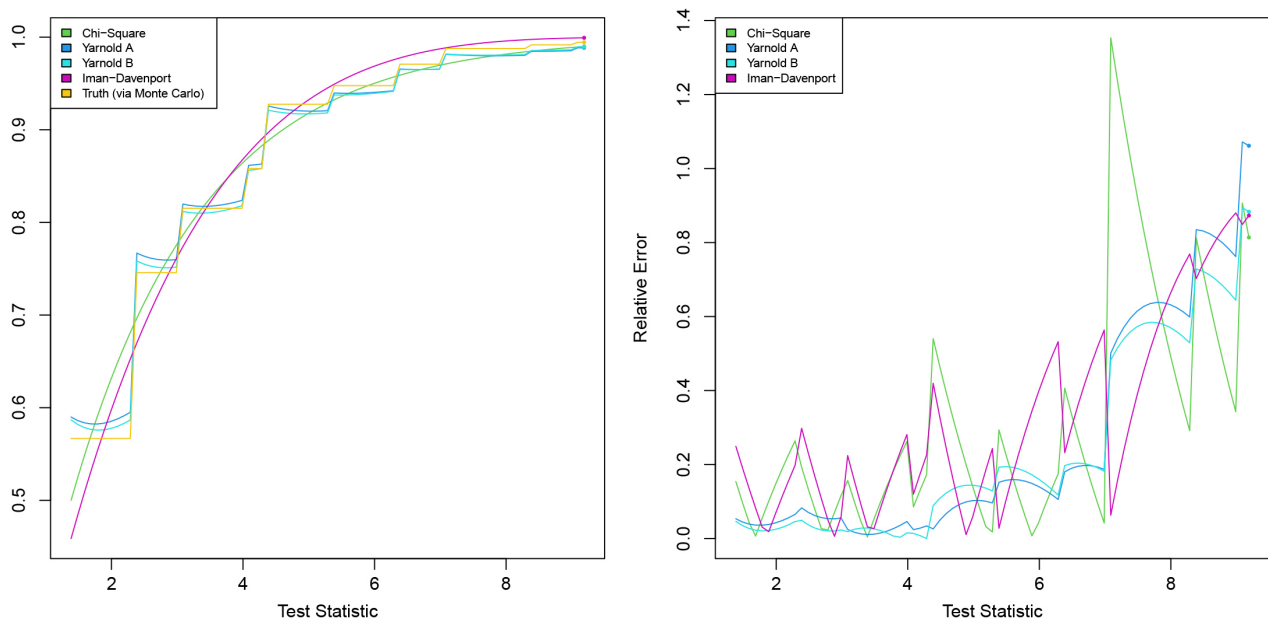
Figure 2. The figure displays the distribution of the Friedman test statistic (left) and the relative error with respect to the distribution of the Friedman test statistic (right) for the case with three groups, six blocks, and one replicate per group.
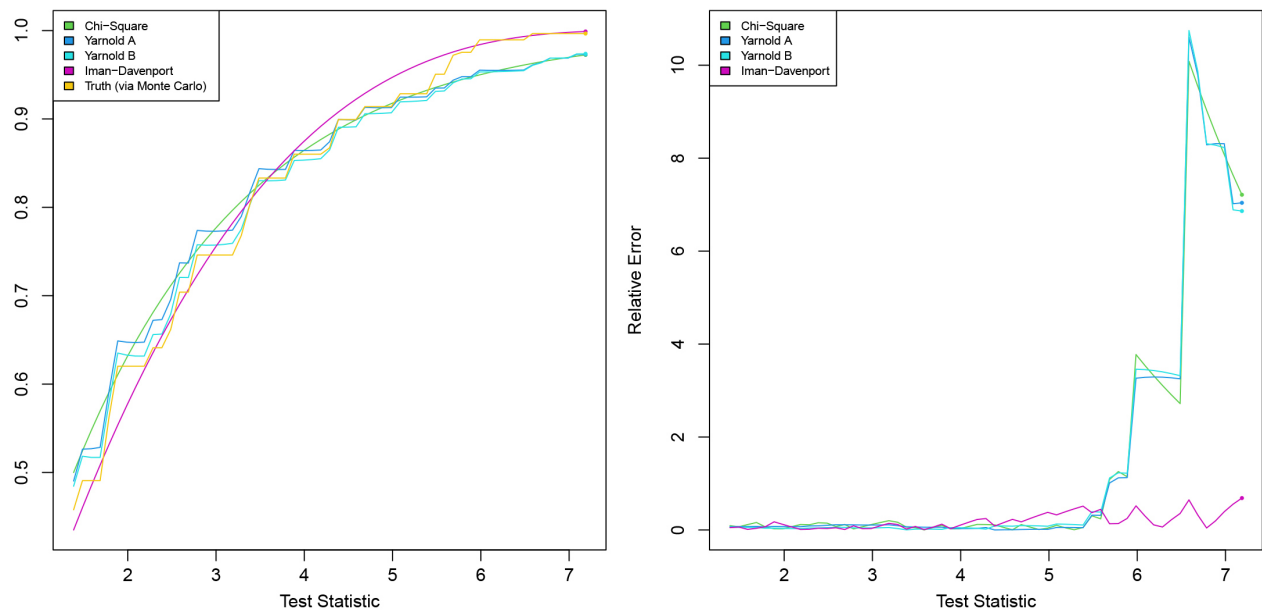

Figure 3. The figure displays the distribution of the Kruskal Wallis test statistic (left) and the relative error with respect to the distribution of the Kruskal Wallis test statistic (right) for the case with three groups, one block, and three replicates per group.

(Mean RE 0.051, SD RE 0.044) and Iman Davenport (Mean RE 0.058, SD RE 0.045) approximations.

Increasing the number of replicates improves the performance of approximations A (Mean RE 0.060, SD RE 0.068) and B (Mean RE 0.057, SD RE 0.057) over the Chi-Square (Mean RE 0.064, SD RE 0.069) approximation.

See an example with three replicates in **Figure 6**.

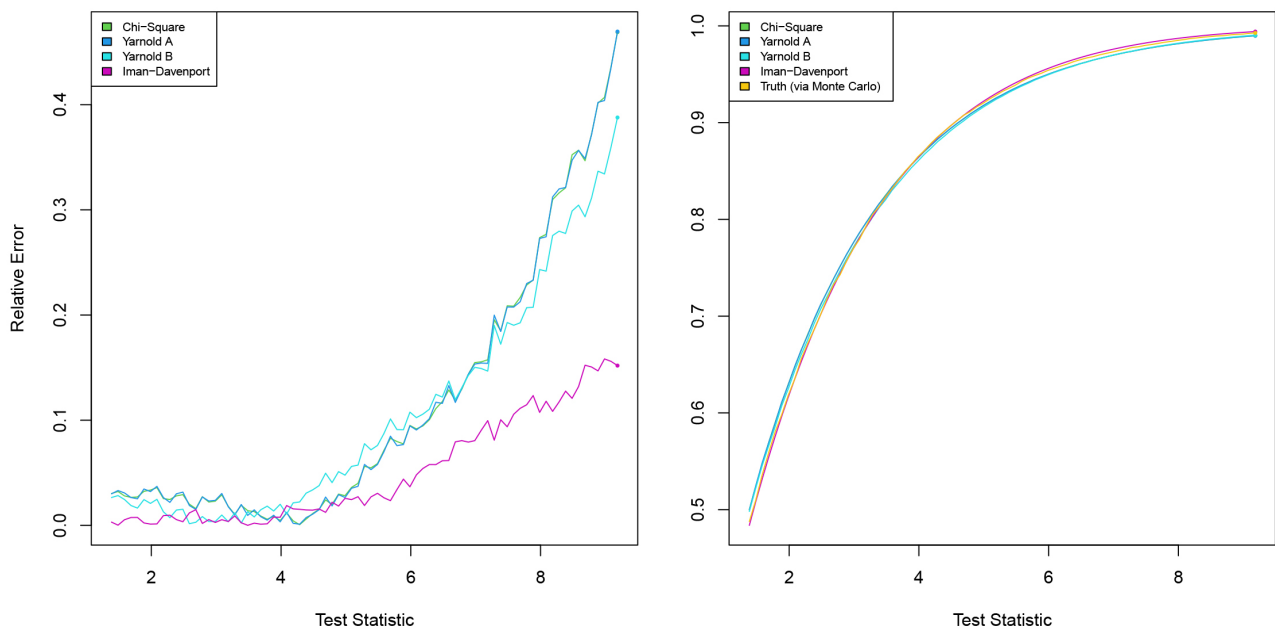The most significant limitation of approximations A and B occurs in the case

Figure 4. The figure displays the distribution of the Kruskal Wallis test statistic (left) and the relative error with respect to the distribution of the Kruskal Wallis test statistic (right) for the case with three groups, one block, and ten replicates per group.
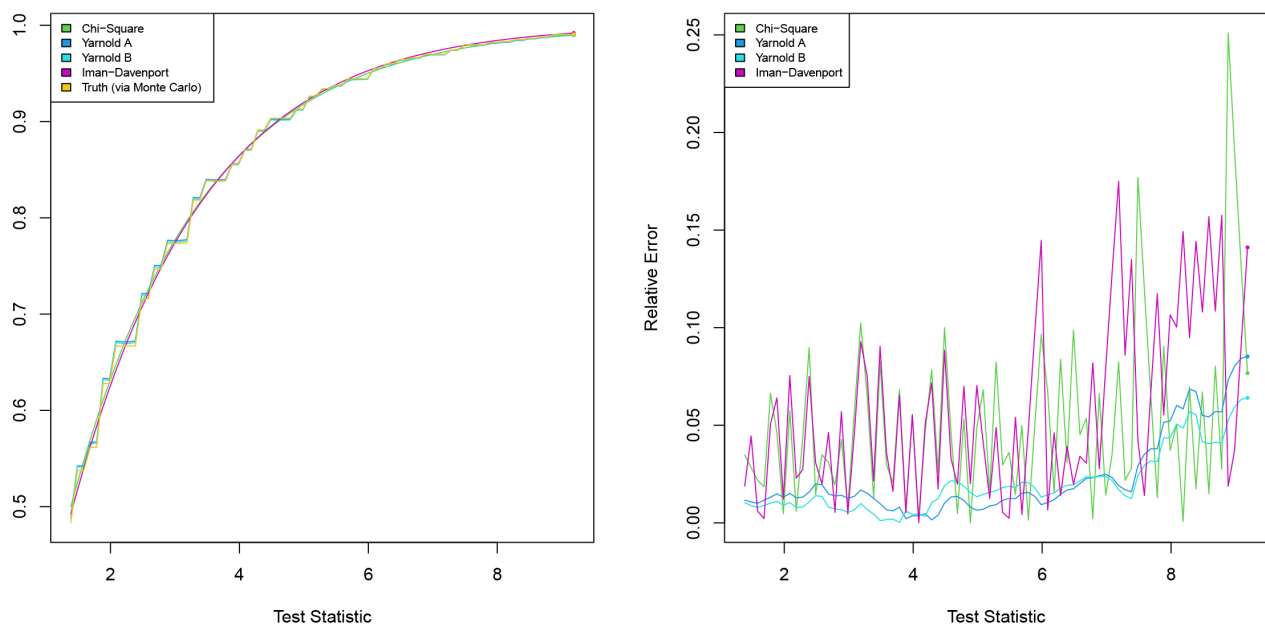

Figure 5. The figure displays the distribution of the Friedman test statistic (left) and the relative error with respect to the distribution of the Friedman test statistic (right) for the case with three groups, thirty blocks, and one replicate per group.

with higher group counts. In these cases, the distribution of the Prentice test statistic exhibits more frequent but smaller discontinuities, and appears more continuous when plotted. Hence, the correction for continuity in the Yarnold A

(Mean RE 0.373, SD RE 0.4223) has a far lesser effect than the cases considered previously. The adjustment for kurtosis in the Yarnold B

(Mean RE 0.3303, SD RE 0.3473) approximation yields a better approximation
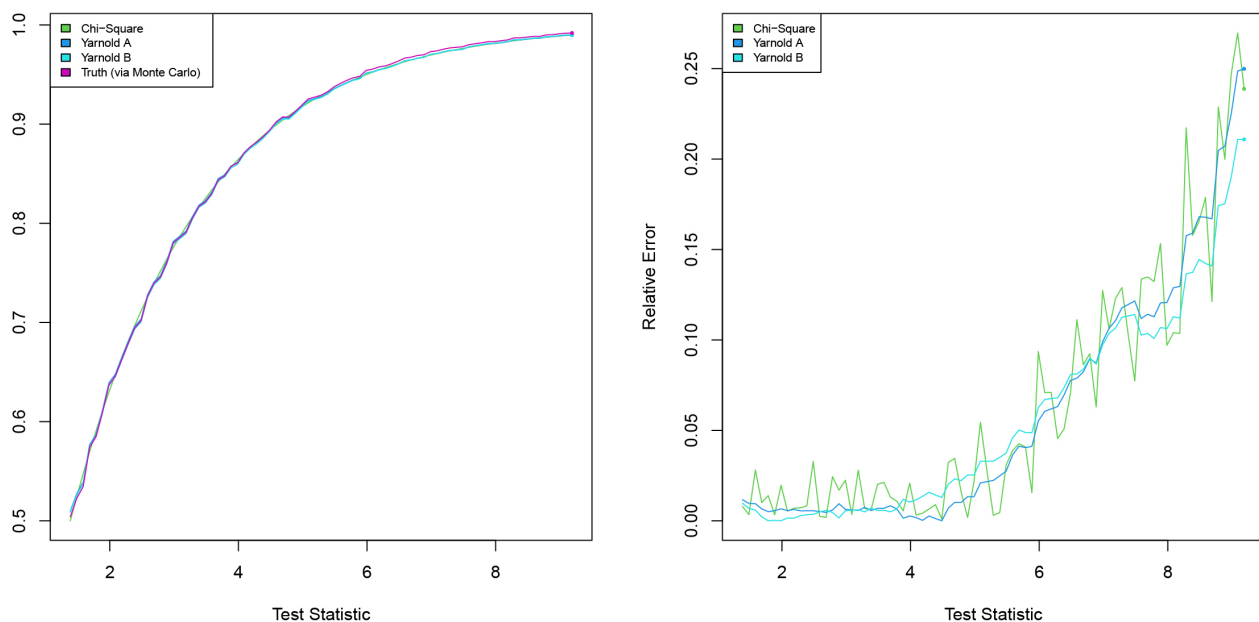
Figure 6. The figure displays the distribution of the Friedman test statistic (left) and the relative error with respect to the distribution of the Friedman test statistic (right) for the case with three groups, six blocks, and three replicates per group.

than the chi-square

(Mean RE 0.373, SD RE 0.422) and Yarnold A approximations in terms of relative error. However, the more significant correction for continuity in the Iman-Davenport approximation

(Mean RE 0.110, SD RE 0.108) yields a much better approximation in terms of relative error than the other approximations. See Figure 7 for an example.

Lastly, we present the effects of an alternative logarithmic scoring system. This results in more frequent discontinuities than in the previous cases considered due to the non-discrete nature of the scores, rendering the correction for continuity minimally effective. Hence, only the first and third terms from approximation B (Mean RE 0.371, SD RE 0.500) were utilized as a comparison to the Chi-Square approximation (Mean RE 0.396, SD RE 0.561).

See Figure 8 for an example.

4. Discussion

Generally, approximation A is at least as good as the Chi-square distribution and approximation B is better than approximation A. This pattern indicates that the correction for kurtosis in approximation B has a greater effect than the correction for continuity in approximations A and B. Even though this pattern holds overall, there are some exceptions where the performance of the Chi-square distribution exceeds that of approximations A and B and when the performance of approximation A exceeds that of approximation B. However, it should be noted that approximation B is most often the best approximation for the tail probability of each distribution.
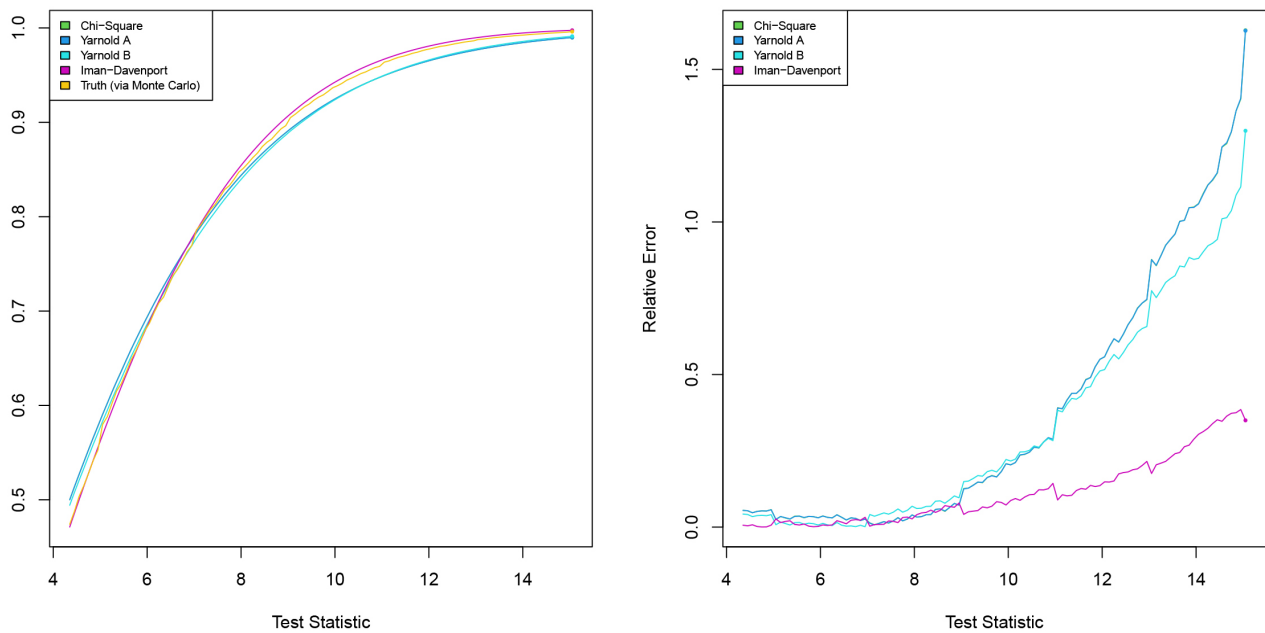
Figure 7. The figure displays the distribution of the Friedman test statistic (left) and the relative error with respect to the distribution of the Friedman test statistic (right) for the case with six groups, six blocks, and one replicate per group.
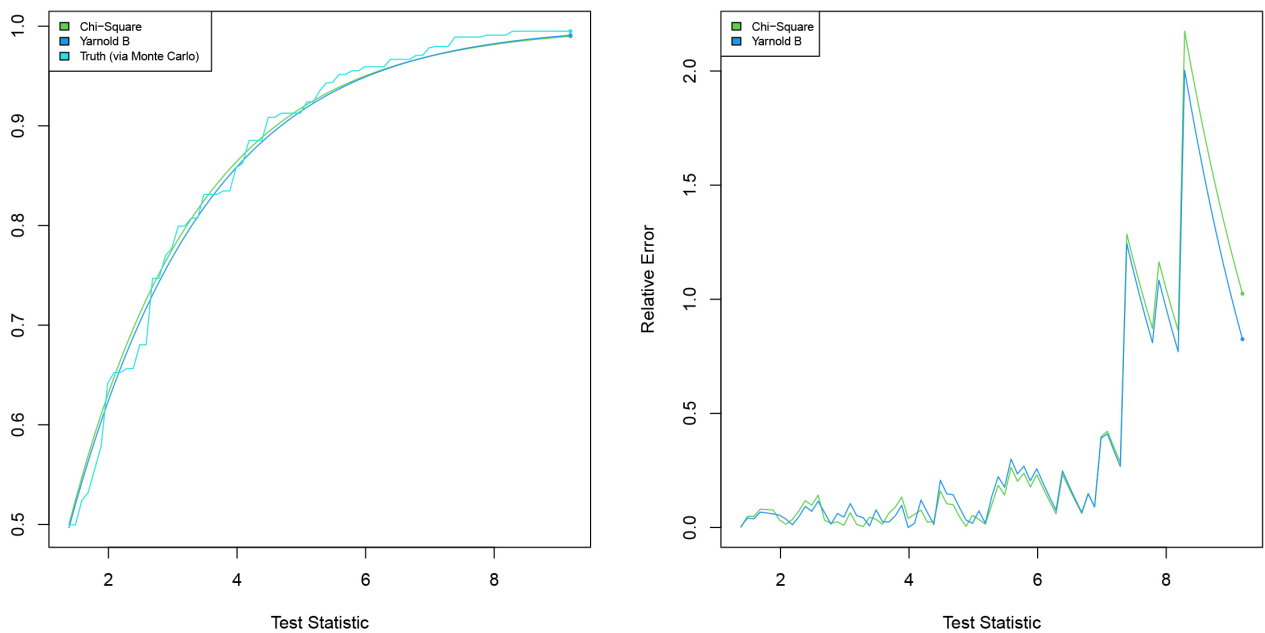

Figure 8. The figure displays the distribution of the Friedman test statistic with logarithmic scoring (left) and the relative error with respect to the distribution of the Friedman test statistic with logarithmic scoring (right) for the case with three groups, six blocks, and one replicate per group.

In cases with one replicate per group, both approximations A and B frequently outperform the Iman-Davenport approximation [8] [9]. However, this does not hold true in all cases and the Iman-Davenport approximation frequently outperforms approximations A and B in cases with high group counts or low block counts. In **Figure 3** which demonstrates the effect of low block counts, some lines are

terminated early to account for the early termination of the Kruskal-Wallis approximation relative to the Chi-square, A, and B approximations. Each terminated line is ended with a bullet point for clarity. In this case, in particular, it is recommended that the Iman-Davenport statistic approximation is used over other approximations, since the high relative error of the Chi-square, A, and B approximations renders them inaccurate approximations to the Kruskal-Wallis test statistic.

With increasing group counts, the relative accuracy of approximations A and B remains unchanged. This is demonstrated by the consistently low relative accuracy of the approximations with low group counts in Figure 2 and higher group counts in Figure 7.

However, the relative accuracy of approximations A and B increases with a high number of blocks, as demonstrated by the example in Figure 5. These effects result from the dependence on the block counts in the standard deviation σ of the Friedman test statistic [3]:

$$\sigma = \sqrt{\frac{p^2 - 1}{12b}}$$

In the formula above, p refers to the number of ranks in the design and b the number of blocks in the design. As shown, the standard deviation of the Friedman test statistic is inversely related to the number of blocks, and as the number of blocks increases, the standard deviation decreases. Therefore, the impact of the correction for continuity in the second term of our approximation decreases, reducing the relative accuracy of both approximations A and B.

The effect of high numbers of replicates is somewhat more significant than that for high numbers of blocks, as demonstrated by the relative error decrease for a modest increase in replicates in 6. With high numbers of replicates, the relative error quantity for all approximations is so small as to deem all approximations equal. Therefore, for computational simplicity, it is recommended that the chi-square approximation is used in these cases since the calculation of $N(nc)$ quickly becomes less efficient as the number of replicates increases in approximations A and B.

Lastly, the use of alternative non-polynomial scoring systems results in sums of scores by treatment that is not supported on a lattice. Hence, the typically discrete distribution is closer to a continuous distribution and the correction for continuity in Yarnold A is not necessary. However, the correction for kurtosis in Yarnold B presents an improvement to the chi-square approximation, as demonstrated by the lower relative error in Figure 8. Also, the $delta_2$ term is non-zero in this case, reflecting the skewness of the underlying score sum distribution due to the dependence of $delta_2$ on the third multivariate cumulant. Comparisons to the Iman-Davenport approximation are not included as the alternative scoring system cannot be applied.

5. Conclusions

We presented an approximation to the Prentice test statistic with corrections for

continuity and kurtosis in approximations A and B [10].

The approximation presents an improvement on the previous Chi-square and Iman-Davenport approximations to the Prentice test statistic. The Yarnold approximation is particularly effective for large block counts with limitations when applied to scenarios with large group counts.

The approximation also presents an improvement in the Chi-square distribution with the use of alternative non-polynomial scoring systems.

Supported

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Prentice, M.J. (1979) On the Problem of m Incomplete Rankings. *Biometrika*, **66**, 167-170. https://doi.org/10.2307/2335259

[2] Kruskal, W.H. and Wallis, W.A. (1952) Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, **47**, 583-621. https://doi.org/10.1080/01621459.1952.10483441

[3] 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. https://doi.org/10.1080/01621459.1937.10503522

[4] Cressie, N. and Read, T.R. (1984) Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **46**, 440-464. https://doi.org/10.1111/j.2517-6161.1984.tb01318.x

[5] Jensen, D. (1977) On Approximating the Distributions of Friedman's χ^2 r and Related Statistics. *Metrika*, **24**, 75-85. https://doi.org/10.1007/BF01893394

[6] Gaunt, R.E., Pickett, A.M. and Reinert, G. (2017) Chi-Square Approximation by Stein's Method with Application to Pearson's Statistic. *Annals of Applied Probability*, **27**, 720-756. https://doi.org/10.1214/16-AAP1213

[7] Gaunt, R.E. and Reinert, G. (2023) Bounds for the Chi-Square Approximation of Friedman's Statistic by Stein's Method. *Bernoulli*, **29**, 2008-2034. https://doi.org/10.3150/22-BEJ1530

[8] Iman, R.L. and Davenport, J.M. (1980) Approximations of the Critical Region of the Friedman Statistic. *Communications in Statistics-Theory and Methods*, **9**, 571-595. https://doi.org/10.1080/03610928008827904

[9] Iman, R.L. and Davenport, J.M. (1976) New Approximations to the Exact Distribution of the Kruskal-Wallis Test Statistic. *Communications in Statistics Theory and Methods*, **5**, 1335-1348. https://doi.org/10.1080/03610927608827446

[10] Yarnold, J.K. (1972) Asymptotic Approximations for the Probability That a Sum of Lattice Random Vectors Lies in a Convex Set. *The Annals of Mathematical Statistics*, **43**, 1566-1580. https://doi.org/10.1214/aoms/1177692389

[11] Esseen, C. (1945) Fourier Analysis of Distribution Functions. A Mathematical Study of the Laplace-Gaussian Law. *Acta Mathematica*, **77**, 1-125. https://doi.org/10.1007/BF02392223

[12] De Leeuw, J. (2012) Multivariate Cumulants in R. https://escholarship.org/content/qt1fw1h53c/qt1fw1h53c_noSplash_8cb15933a039988ef5b788a1b4ef1b38.pdf?t=mkq6eg

[13] Kolassa, J. (2020) An Introduction to Nonparametric Statistics. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9780429202759

[14] Minitab, L.L.C. (2023) Example of Friedman Test. https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistics/nonparametrics/how-to/friedman-test/before-you-start/example/

117