

Calculation of Two-Tailed Exact Probability in the Wald-Wolfowitz One-Sample Runs Test

José Moral De La Rubia

School of Psychology, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, México
Email: jose.morald@uanl.edu.mx

How to cite this paper: Moral De La Rubia, J. (2024) Calculation of Two-Tailed Exact Probability in the Wald-Wolfowitz One-Sample Runs Test. *Journal of Data Analysis and Information Processing*, 12, 89-114.
<https://doi.org/10.4236/jdaip.2024.121006>

Received: December 19, 2023

Accepted: February 19, 2024

Published: February 22, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The objectives of this paper are to demonstrate the algorithms employed by three statistical software programs (R, Real Statistics using Excel, and SPSS) for calculating the exact two-tailed probability of the Wald-Wolfowitz one-sample runs test for randomness, to present a novel approach for computing this probability, and to compare the four procedures by generating samples of 10 and 11 data points, varying the parameters n_0 (number of zeros) and n_1 (number of ones), as well as the number of runs. Fifty-nine samples are created to replicate the behavior of the distribution of the number of runs with 10 and 11 data points. The exact two-tailed probabilities for the four procedures were compared using Friedman's test. Given the significant difference in central tendency, post-hoc comparisons were conducted using Conover's test with Benjamini-Yekutieli correction. It is concluded that the procedures of Real Statistics using Excel and R exhibit some inadequacies in the calculation of the exact two-tailed probability, whereas the new proposal and the SPSS procedure are deemed more suitable. The proposed robust algorithm has a more transparent rationale than the SPSS one, albeit being somewhat more conservative. We recommend its implementation for this test and its application to others, such as the binomial and sign test.

Keywords

Randomness, Nonparametric Test, Exact Probability, Small Samples, Quantiles

1. Introduction

The one-sample randomization test was conceived in 1943 by the Hungarian-born mathematician Abraham Wald (1902-1950) and the Polish-born mathematician

Jacob Wolfowitz (1910-1981) [1]. Both authors had immigrated to the United States and crossed paths in New York City. At the time of developing the runs test, Wald held a professorship at Columbia University, while Wolfowitz was pursuing his doctorate at New York University. They fostered a friendship and collaboration on various statistical endeavors until Wald's untimely death in a plane crash during a trip to India. He had been invited by the Indian government to deliver a series of lectures on econometrics and applied statistics.

It's noteworthy that the runs test, aimed at determining whether two random samples have been drawn from the same population, preceded the test of randomness for a single sample [2]. In both tests, the statistical hypotheses are two-tailed. Both tests count the number of runs (of zeros and ones), and these test statistics have the same distribution. When the sample is small, it is necessary to calculate the exact two-tailed probability, whereas for a large sample, the asymptotic normal probability is used. Because the distribution of the number of runs is not symmetric, unless its two parameters have the same value, the calculation of the two-tailed exact probability is complicated and presents several approaches.

This study has three objectives to: 1) illustrate the algorithms utilized by three statistical computing programs, namely R [3], Real Statistics using Excel (RSUE) [4], and Statistical Package for the Social Sciences (SPSS) [5] [6] [7], for calculating the exact two-tailed probability of the Wald-Wolfowitz runs test [1]; 2) propose a new method for calculating this probability based on quantiles (left side) and complementary quantiles (right side), establishing the two sides of the number of runs distribution with its median; and 3) compare the four procedures using samples of 10 data points (even n) and 11 data points (odd n), varying the parameters n_0 and n_1 , as well as the number of runs (r).

Two open statistical computing programs of wide diffusion and development, R and RSUE, and one closed one, SPSS, which is particularly prevalent in psychology and related sciences [8], were selected for this study. Since the calculation of the two-tailed exact probability is necessary to test randomness with small samples using the Wald-Wolfowitz test, the study is justified in that the RSUE procedure is inadequate because it assumes symmetry and makes the calculation analogous to the normal distribution, which is symmetrical. However, the distribution of the number of runs is symmetrical only when its two parameters, n_0 (number of zeros) and n_1 (number of ones), have equal values.

The R program employs a straightforward procedure of doubling the probability to one tail. However, it is not entirely suitable for various situations of asymmetry in the distribution of the number of runs. Even when the one-tailed probability is greater than 0.5, it results in a two-tailed probability greater than 1. Additionally, it defines the two sides of the distribution based on mathematical expectation, while using the median would be more appropriate [9].

On the other hand, the SPSS program utilizes the mathematical expectation of the number of runs as its axis of symmetry, and the absolute distance between r

and the mathematical expectation allows it to establish the number of runs that correspond on the other side of the distribution ($r_{other-tail}$). The sum of the one-tailed probabilities of r and $r_{other-tail}$ gives the two-tailed probability. In comparison to the two preceding methods, this approach is more suitable. However, opting for the median as the axis of symmetry in various situations of skewness in the distribution of the number of runs appears more appropriate than using the mathematical expectation or arithmetic mean [9].

The proposed new procedure introduces a robust algorithm that adapts to different conditions of skewness in the distribution of the number of runs. This algorithm is based on quantiles (left side) and complementary quantiles (right side), with the median serving as the axis of symmetry.

The article starts by presenting the statistical hypotheses and test assumptions. It then proceeds to explain the derivation of the testing statistic (number of runs) and its distribution, highlighting differences among the R program, RSUE, and SPSS. A proposal is made for calculating the two-tailed exact probability. Next, 59 dichotomous samples with 10 and 11 data points are generated, varying parameters n_0 and n_1 , and the number of runs (2 to the maximum) to observe the behavior of the four algorithms and make comparisons. Finally, conclusions are drawn, and suggestions are provided.

2. Statistical Hypotheses and Assumptions

When applying the one-sample runs test to assess the assumption of randomness or independence within a sequence of sample data, the formulation of statistical hypotheses is conducted in a two-tailed manner [10]. The SPSS program uniquely provides this formulation as an option [5].

The RSUE program offers two alternatives: two-tailed or one-tailed. Opting for the one-tailed alternative, the program disregards directionality, performing calculations in absolute value [4].

The R program presents three options: two-tailed (indicating an alternative hypothesis of non-randomness), left-tailed (enabling the testing of the null hypothesis of randomness against an alternative hypothesis of a downward trend), and right-tailed (allowing the testing of the null hypothesis of randomness against an alternative hypothesis of an upward trend) [3]. The default option in all three programs is two-tailed.

The test only necessitates a sample of n data points from a variable X , whether it be a qualitative, ordinal, or quantitative variable, drawn from a population. To apply the test, the data is dichotomized based on a criterion, unless it is already dichotomous, resulting in $X \rightarrow D = \{0, 1\}$. The count of runs of zeros (n_0) and ones (n_1) is then determined. A run is defined as a continuous succession of the same element (0 or 1) or element class ($X < \text{criterion}$ or $X \geq \text{criterion}$ with SPSS, and $X < \text{criterion}$ or $X > \text{criterion}$ with RSUE and R) in the sequence of n extractions. When a different element or element class is extracted, the run is considered changed.

3. Test Statistic

The test statistic is the number of runs, denoted by R (variable) and r (value). We commence by dichotomizing the n sample data of the variable X , unless X is already a dichotomous qualitative variable. In the case of a polychotomous qualitative variable, it can be dichotomized by the mode (mo), if it exists and is unique. For an ordinal variable, dichotomization can be carried out by the median (mdn). For a quantitative variable, dichotomization can be performed either by the median (mdn) or by the arithmetic mean (m), although using the median is more customary, which appears as the default option in all three programs. Alternative criteria may be considered, such as a quantile other than the median, a type of mean other than the average, or the estimation of the mode in a continuous variable [11]. Let x be a random sample of size n from a variable X : $x = \{x_i\}_{i=1}^n = \{x_1, x_2, \dots, x_n\} \subseteq X$.

Criteria to dichotomize the n sample data of the variable X :

- $mo(x)$ = mode or most frequent value among the n sample data (Equation (1)). To choose this criterion requires unimodality, that is, that the modal value is unique.

$$\begin{array}{cc} \frac{x_i}{x_1} & \frac{n_i}{n_{x_1}} \\ \vdots & \vdots \\ \frac{x_k}{\Sigma} & \frac{n_{x_k}}{n} \end{array} \tag{1}$$

$$mo(x) = \left\{ x_i \mid \max \left(\left\{ n_{x_i} \right\}_{i=1}^k \right) \right\}$$

- $mdn(x)$ = median or central value of the n data sorted in ascending order (Equation (2)).

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$mdn(x) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is even} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is odd} \end{cases} \tag{2}$$

- $m(x)$ = arithmetic mean or average value obtained by summing the n sample data and dividing by the total number of data points (Equation (3)).

$$m(x) = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} \tag{3}$$

By dichotomizing the n sample data of X , two groups are created: the group of zeros and the group of ones (Equation (4)).

$$D = \{0, 1\}$$

$$d = \{d_i\}_{i=1}^n = \{d_1, d_2, \dots, d_n\} \tag{4}$$

$$d_i = \begin{cases} 0 & \text{if } x_i < \text{criterion} : mo(x), mdn(x), \text{ or } \bar{x} \\ 1 & \text{if } x_i \geq \text{criterion} : mo(x), mdn(x), \text{ or } \bar{x} \end{cases}$$

In addition, the following statistics are calculated:

- n_0 = number of zeros in the dichotomized sample.
- n_1 = number of ones in the dichotomized sample.
- $n = n_0 + n_1$. These two groups are independent and mutually exclusive. Furthermore, the classification of participants into these groups is exhaustive, ensuring that no cases are lost, and each case is assigned to a group. Therefore, the sum of the number of elements in both groups corresponds to the sample size.
- r = number of runs. A run is a sequence of zeros or ones in a row. Therefore, the number of changes in the sequence of the n values of D (in the order of collection of the corresponding n sample data of X) is counted and one is added to obtain the number of runs.

This is the classical procedure of Wald and Wolfowitz [1], applied by the SPSS program [5]. However, R [3] and RSUE [4] use a tie correction method that involves removing values equal to the criterion, causing the sum of n_0 and n_1 to be less than or equal to the sample size.

Once the number of runs (r) is determined, the probability value is calculated, either exact or asymptotic, depending on the sum of n_0 and n_1 . The decision is then made based on this probability value. The decision can also be determined by the position of the number of runs relative to critical values or quantiles, which can be either exact or asymptotic depending on the sum of n_0 and n_1 .

4. Sampling Distribution and Decision with Small Samples

When n_0 or n_1 is less than or equal to 20, the exact probability of the number of runs is calculated. For even runs, use the formula in Equation (5); for odd runs, employ the formula in Equation (6) [2].

$$P(R = r) = \frac{2 \times \binom{n_0 - 1}{\frac{r}{2} - 1} \times \binom{n_1 - 1}{\frac{r}{2} - 1}}{\binom{n}{n_0}} \tag{5}$$

$$P(R = r) = \frac{\binom{n_0 - 1}{\frac{r+1}{2} - 1} \times \binom{n_1 - 1}{\frac{r+1}{2} - 2} + \binom{n_0 - 1}{\frac{r+1}{2} - 2} \times \binom{n_1 - 1}{\frac{r+1}{2} - 1}}{\binom{n}{n_0}} \tag{6}$$

The number of runs has a minimum and a maximum. The minimum is 2. This minimum is achieved when all values less than the cutoff point (zeros) first appear, followed by all values greater than or equal to the cutoff point (ones), or vice versa. The upper limit occurs when these two values of the dichotomous or dichotomized variable (by median, mean, mode, or other value) alternate in their collection sequence. When $n_0 = n_1$, the maximum is n . When $n_0 \neq n_1$, the

maximum is $2 \times \min(n_0, n_1) + 1$. Both extremes are examples of nonrandom sequences.

By summing the individual point probabilities progressively from 2 to n when $n_0 = n_1$, or from $2 \times \min(n_0, n_1) + 1$ when $n_0 \neq n_1$, we obtain the cumulative distribution function, $F(R|n_0, n_1)$, of the Wald-Wolfowitz distribution for the number of runs [1].

Calculating the exact one-tailed probability involves defining the two sides (left and right) of the distribution of the number of runs, for example, by its median or its mathematical expectation, and determining on which side r (the number of runs in the sample) is located. If r is on the left side, the exact point probabilities from 2 to r are added. If r is on the right side, the exact point probabilities from r to the maximum are summed. The calculation of the bilateral exact probability is more complex and varies from one statistical computing program to another. Below is its computation with SPSS, R program, RSUE, and the new proposal.

4.1. Exact Probability Using SPSS

Mehta and Patel [7] clarify that the SPSS program calculates the two-tailed probability using the definition in Equation (7). In this equation, R represents the variable or distribution of the number of runs with parameters n_0 and n_1 , r is the number of runs in the sample, and $E(R)$ is the mathematical expectation of the distribution.

$$P_{2\text{-colas}} = P\left(|R - E(R|n_0, n_1)| \geq |r - E(R|n_0, n_1)|\right) \quad (7)$$

Following Equation (7), the sides of the distribution are defined by the change in sequence of $|r - E(R)|$, with the values of r ordered from 2 (minimum) to maximum. The downward, or negative sequence with no absolute value, occurs when $r \leq E(R)$ or $r - E(R) \geq 0$, forming the left side. The ascending, or positive sequence with no absolute value, occurs when $r > E(R)$ or $r - E(R) < 0$, forming the right-hand side. Therefore, the two sides of the distribution are defined based on the mathematical expectation of R (Equation (8)).

$$E(R|n_0, n_1) = \frac{2 \times n_0 \times n_1}{n_0 + n_1} + 1 \quad (8)$$

Next, the one-sided probability corresponding to the side on which r is located is calculated. Let r be the number of runs in a sample of size n from a variable X with n_0 sample data less than the criterion (median, arithmetic mean, mode, or other statistic of the n sample data of X) and n_1 sample data greater than or equal to the criterion, where $n_0 + n_1 = n$. When r is located on the left side, the point probabilities from 2 to r are summed (Equation (9)). When r is located on the right side, the point probabilities from r to the maximum ($\max(R) = n$ when $n_0 = n_1$ or $2 \times \min(n_0, n_1) + 1$ when $n_0 \neq n_1$) are summed (Equation (10)).

$$r \leq E(R | n_0, n_1) \quad p_{left\ tail} = P(R \leq r) = \sum_{i=2}^r P(R = i) \quad (9)$$

$$r > E(R | n_0, n_1) \quad p_{right\ tail} = P(R \geq r) = \sum_{i=r}^{\max(R)} P(R = i) \quad (10)$$

To each value of r corresponds a single value $|r - E(R)|$. If $n_0 = n_1$, each value of $|r - E(R)|$ corresponds to two values of r , one on each side of the distribution. However, if $n_0 \neq n_1$, it corresponds to a single value. To obtain the two-tailed probability, the one-tailed probability of r is added to the probability toward the other tail of the value of R that is equal (when $n_0 = n_1$) or immediately greater (when $n_0 \neq n_1$) than $|r - E(R)|$ in that tail. If there is none, the probability to the other tail is 0. In Equation (11), the algorithm for calculating the exact two-tailed probability for a value of r on the left side is shown; such value is denoted by r_{LS} and the value of R corresponding to the right side is denoted by r_{RS} . In Equation (12), it is shown for a value of r on the right side, which is denoted by r_{RS} and the value of R corresponding to the left side by r_{LS} .

$$r_{LS} \leq E(R | n_0, n_1)$$

$$p_{left\ tail} = P(R \leq r_{LS} \rightarrow |r_{LS} - E(R)|) = \sum_{i=2}^{r_{LS}} P(R = i) \quad (11)$$

$$p_{other\ tail} = P(R \geq r_{RS} \rightarrow |r_{LS} - E(R)|) = \sum_{i=r_{RS}}^{\max(R)} P(R = i)$$

$$p_{2\ tails} = P(|R - E(R)| \geq |r_{LS} - E(R)|) = p_{left\ tail} + p_{other\ tail}$$

$$r_{RS} > E(R | n_0, n_1)$$

$$p_{right\ tail} = P(R \geq r_{RS}) = \sum_{i=r_{RS}}^{\max(R)} P(R = i) \quad (12)$$

$$p_{other\ tail} = P(R \leq r_{LS} \rightarrow |r_{RS} - E(R)|) = \sum_{i=2}^{r_{LS}} P(R = i)$$

$$p_{2\ tails} = P(|R - E(R)| \geq |r_{RS} - E(R)|) = p_{right\ tail} + p_{other\ tail}$$

See the distribution of the number of runs (R) with parameters: $n_0 = 7$ (values less than the criterion: median) and $n_1 = 2$ (values greater than or equal to the criterion), where $n = n_0 + n_1 = 7 + 2 = 9$ (sample size). In this distribution taken as an example, the minimum value is: $min(R | n_0 = 7, n_1 = 2) = 2$, the maximum value is: $max(R | n_0 = 7, n_1 = 2) = 2 \times min(n_0, n_1) + 1 = 5$, and the expected value is: $E(R | n_0 = 7, n_1 = 2) = 1 + (2 \times n_0 \times n_1) / n = 4.111$. If a significance level of 0.1 is adopted, taking into account the small sample size ($n = 9$), the null hypothesis of randomness is rejected with a number of runs of 2. However, with three other numbers of runs ($r = 3, 4, \text{ or } 5$), the null hypothesis holds. **Table 1** shows the exact point probabilities, calculated using Equations (5) and (6), as well as the one-tailed probability, obtained using Equations (9) and (10), and the other-tailed and two-tailed probabilities, computed using Equations (11) and (12).

4.2. Exact Probability Using R Program

The R program calculates the two-tailed exact probability by simply doubling the

Table 1. One- or two-tailed probability for a distribution of the number of runs with parameters $n_0 = 7$ and $n_1 = 2$ using SPSS.

r	$P(R = r)$	$ r - E(R) $	Side	One-tailed p -value	$r_{other\ side}$	Other-tailed p -value	Two-tailed p -value
2	0.056	2.111	Left	0.056	0	0	0.056
3	0.194	1.111	Left	0.250	0	0	0.250
4	0.333	0.111	Left	0.583	5	0.417	1
5	0.417	0.889	Right	0.417	3	0.250	0.667

Note. r = number of runs, $P(R = r)$ = exact point probability, $|r - E(R)|$ = absolute difference between the number of runs and the mathematical expectation of the $R|n_0, n_1$ distribution, Side = Left when $r \leq E(R) = 4.111$ and Right when $r > E(R)$, One-tailed p -value = sum of the exact point probabilities from 2 to r on the left side and from r to the maximum on the right side, $r_{other\ side}$ = value of r on the other side of the distribution that corresponds to a value immediately greater than $|r - E(R)|$, other-tail p -value = one-tailed probability value of $r_{other\ side}$ y Two-tailed p -value = sum of the one-tailed and the other-tailed probability values.

one-tailed exact probability [3]. It's important to note that a value greater than 1 may appear (Equation (13)). The distribution is divided into two parts based on the mathematical expectation or arithmetic mean of the number of runs (Equation (8)). Values of R less than or equal to its mean constitute the left-hand side of the distribution, while values of R greater than the mean form the right-hand side. If r is less than or equal to $E(R)$, the exact probability for a tail (left tail) is obtained by adding the exact point probabilities from 2 (minimum value of R) to r (Equation (13)). If r is greater than $E(R)$, the one-tailed exact probability (right tail) is calculated by summing the pointwise exact probabilities from r to the maximum of R : $n_0 + n_1$ when $n_0 = n_1 \leq n$ or $2 \times \min(n_0, n_1) + 1$ when $n_0 \neq n_1$ (Equation (14)).

$$p_{2\ tails} = 2 \times p_{1\ tail} \tag{13}$$

$$p_{1\ tail} = \begin{cases} \sum_{i=2}^r P(R = i) & r \leq E(R) \\ \sum_{i=r}^{\max(R)} P(R = i) & r > E(R) \end{cases} \tag{14}$$

The script for executing the runs test using the R program is displayed. Hypotheses can be formulated as two-tailed (alternative = "two.sided"), left-tailed (alternative = "left.sided"), or right-tailed (alternative = "right.sided"). The dichotomization criterion for the sample is specified in 'threshold,' with the default being the median. If the data are dichotomous with coding 0 and 1, the value of 0.5 can serve as the dichotomization criterion. The probability value calculation can be exact (p value = "exact") or asymptotic (p value = "default"). By default, it provides the asymptotic probability, unless $n_0 + n_1$ is less than 10. A two-dimensional plot can be included (plot = TRUE) or excluded (plot = FALSE). This plot allows visualization of the sequence of the n sample data. On the abscissa axis, the random order of the sample data is displayed, and on the ordinate axis, the value of the data on the X scale.


```

#Vector of sample data.
x <- c(5.59, 4.13, 7.18, 4.96, 5.59, 8.42, 4.71, 6.13, 4.2)
library(randtests)
#Criterion: Median.
mdn<- median(x)
cat("median =",mdn,"\n")
runs.test(x, alternative = "two.sided", threshold = mdn, pvalue = 'exact', plot = TRUE)
#Criterion: Arithmetic mean.
m <- mean(x)
cat("mean =",m,"\n")
runs.test(x, alternative = "two.sided", threshold = m, pvalue = 'exact', plot = TRUE)
#Criterion: Mode for a continuous variable using kernel estimation with the Epanechnikov's
parabolic kernel and the Sheather-Jones bandwidth (Poncet, 2022).
library(modeest)
mo <- mlv(x, method = "parzen", bw = "SJ", kernel = "epanechnikov")
cat("Parzen's mode =",mo,"\n")
runs.test(x, alternative = "two.sided", threshold = mo, pvalue = 'exact', plot = TRUE)
#Criterion: Mode for a discrete variable using the maximum frequency value.
library(modeest)
mod <- mlv(x, method = "mfv")
cat("mode =",mod,"\n")
runs.test(x, alternative = "two.sided", threshold = mod, pvalue = 'exact', plot = TRUE)

```

In **Table 2**, the previous example is revisited. Using the R program, the null hypothesis of randomness holds with the four values of the distribution of the number of runs with parameters: $n_0 = 7$ (values less than the criterion: median) and $n_1 = 2$ (values greater than or equal to the criterion) in a two-tailed test, as corresponds to this type of hypothesis, and a level of significance of 0.01 considering the small sample size ($n = n_0 + n_1 = 9$).

4.3. Exact Probability Using Real Statistics Using Excel

Like the R program, the RSUE package also has a simple procedure for calculating exact two-tailed probabilities [4]. The two parts of the distribution are defined by the median of the number of runs or the value of R with a cumulative probability of at least 50%. If the number of runs is less than the median of R, the one-sided probability is the sum of the point probabilities from 2 to r , and the two-sided probability is twice this sum (Equation (15)).

$$r < Mdn(R | n_0, n_1) \Rightarrow P(R \leq Mdn(R | n_0, n_1)) \geq 0.5$$

$$p_{1tail} = \sum_{i=2}^r P(R=i) \quad (15)$$

$$p_{2tails} = 2 \times \sum_{i=2}^r P(R=i)$$

If the number of runs is greater than or equal to the median of R, the one-tailed probability is the complement of the sum from 2 to r , and the two-tailed probability is twice that complement (Equation (16)).

Table 2. One- or two-tailed probability for a distribution of the number of runs with parameters $n_0 = 7$ and $n_1 = 2$ using R program.

r	$P(R = r)$	Side	Alternative hypothesis	One-tailed p -value	Alternative hypothesis	Two-tailed p -value
2	0.056	Left	Trend	0.0556	Non-randomness	0.1111
3	0.194	Left	Trend	0.25	Non-randomness	0.5
4	0.333	Left	Trend	0.5833	Non-randomness	1.167
5	0.417	Right	First-order negative	0.4167	Non-randomness	0.8333

Note. r = number of runs, $P(R = r)$ = exact point probability, Side = Left when $r \leq E(R) = 4.111$ and Right when $r > E(R)$, One-tailed p -value = sum of the exact point probabilities from 2 to r on the left side and from r to the maximum on the right side, and Two-tailed p -value = doubling of one-tailed probability value.

$$\begin{aligned}
 r &\geq Mdn(R | n_0, n_1) \\
 p_{1tail} &= 1 - \sum_{i=2}^r P(R = i) \\
 p_{2tails} &= 2 \times \left(1 - \sum_{i=r}^n P(R = i) \right)
 \end{aligned}
 \tag{16}$$

In **Table 3**, the example seen previously with the SPSS and R programs is revisited. The null hypothesis of randomness is rejected with a value of r equal to 5, even though the point probability is not zero, $P(R = 5) = 0.417$, and this hypothesis holds when tested using SPSS and R program. This result is because the value of r equal to 5 accumulates 100% of the distribution, $P(R \leq 5) = 1$, so the complement of its cumulative probability is 0 (one-tailed probability), and the duplicate complement is 0 (two-tailed probability).

4.4. New Proposal for the Calculation of the Exact Two-Tailed Probability

The exact two-tailed probability is obtained by adding the one-sided probability of the side where r is located and the probability corresponding to the other tail. Taking the median as the axis of symmetry, a simple procedure to find the probability corresponding to the other tail is through the quantile order when r is located on the left side and the complementary quantile order when r is situated on the right side.

We begin by calculating the median of R to define the left and right sides of the distribution. The median of R is the value that accumulates at least 50% of the probability mass of the Wald-Wolfowitz distribution of the number of runs [2]. Both the probabilities on the left side of the distribution, from 2 to r , and the probabilities on the right side, from r to the maximum: n ($n_0 = n_1$) or $2 \times \min(n_0, n_1) + 1$ ($n_0 \neq n_1$), are calculated. Additionally, the mode of R (the r value with the highest probability) is computed, and Bickel’s robust skewness measure based on the mode is calculated. This measure is obtained by subtracting from one twice

Table 3. One- or two-tailed probability for a distribution of the number of runs with parameters $n_0 = 7$ and $n_1 = 2$ using real statistics using excel.

r	$P(R = r)$	$P(R \leq r)$	Side	One-tailed p -value	Two-tailed p -value
2	0.056	0.056	Left	0.056	0.111
3	0.194	0.250	Left	0.250	0.5
4	0.333	0.583	Right	0.417	0.833
5	0.417	1	Right	0	0

Note. r = number of runs, $P(R = r)$ = exact point probability, $P(R \leq r)$ = exact cumulative exact probability, Side = Left when $r < Mdn(R) = 4$ and Right when $r \geq Mdn(R)$, One-tailed p -value = sum of the exact point probabilities from 2 to r on the left side and from r to the maximum on the right side, and Two-tailed p -value = doubling of one-tailed probability value.

the cumulative distribution function evaluated at the mode [12]. Denoted by A_B , it varies from -1 to 1 , where 0 indicates symmetry, negative values show left-tailed skewness, and positive values exhibit right-tailed skewness (Equation (17)).

$$A_B = 1 - 2 \times F_x(Mo) \tag{17}$$

The distribution of the number of runs varies from symmetry ($A_B = 0$) when $n_0 = n_1$, where $n_0 + n_1 = n$, to extreme negative skewness ($A_B = -1$) when $n_0 = 1$ and $n_1 = n - 1$ or $n_0 = n - 1$ and $n_1 = 1$ (Figure 1 and Figure 2). With non-extreme skewness values ($0 \geq A_B > -1$), if r is less than the median of R, we calculate the probability that accumulates from 2 (minimum value) to r or the left-tailed probability, representing the order of the quantile. The value with a right-tailed probability equal to or immediately greater is sought, and this probability is the one corresponding to the other tail. The sum of both probabilities provides the two-sided probability. If there is none equal or greater, the one-sided probability is doubled, resulting in the two-tailed probability.

Similarly, if the number of runs is greater than the median of R, we sum the probabilities from r to $\max(R)$ or the right-tailed probability, representing the order of the complementary quantile. The probability to the left tail equal or immediately greater is sought, and this probability is the one corresponds to the other tail. The two-sided probability is the sum of both probabilities. If there is neither, the one-sided probability is doubled, resulting in the two-tailed probability.

If r is the median, the other-tailed (right) probability is its complement, and the two-tailed probability is the sum of both probabilities, yielding a value of 1.

With an extreme skewness value of $A_B = -1$, where the right tail does not exist or one value dominates all the probability, we proceed in the same manner as with non-extreme skewness to calculate the two-tailed probability of the median, which is 1. We double the one-sided probability of the r value that defines the shortened tail to obtain its two-tailed probability. If this doubling exceeds unity,

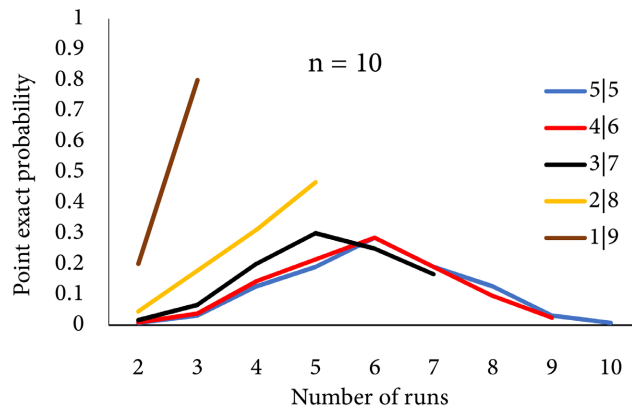


Figure 1. Frequency polygons for the distribution of the number of runs with parameters n_0 and n_1 , where $n_0 + n_1 = 10$ (even).

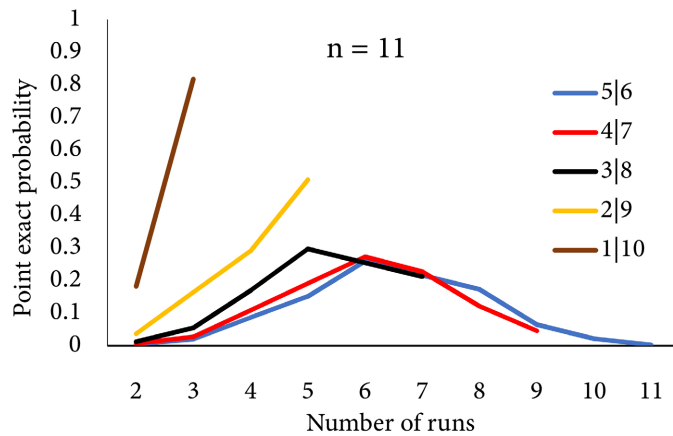


Figure 2. Frequency polygons for the distribution of the number of runs with parameters n_0 and n_1 , where $n_0 + n_1 = 11$ (odd).

the two-tailed probability is given a value of 1. The one-tailed probabilities of the values of the elongated part of the distribution are taken as their two-tailed probabilities, considering that the probabilities of the other tail are zero.

Table 4 shows the exact two-tailed probabilities yielded by the newly proposed algorithm following the previously applied example. Results closely align with SPSS. At a significance level of 10%, the null hypothesis is rejected when $r = 2$. When $r = 5$, the null hypothesis is maintained, and the two-tailed exact probability is slightly higher than that provided by SPSS (0.833 versus 0.833), which is the only difference with SPSS.

4.5. Decision Based on Probability Value and Critical Value

In determining the randomness of the sample data sequence, the decision relies on the bilateral significance. If the two-tailed probability value equals or exceeds the significance level α , the null hypothesis of randomness is upheld. If it falls below α , the hypothesis is rejected. Typically, the significance level is set at 0.05. However, for very small samples ($n \leq 10$), it might be raised to 0.1 to counter-balance the conservatism of the test associated with the null hypothesis.

Table 4. One- or two-tailed probability for a distribution of the number of runs with parameters $n_0 = 7$ and $n_1 = 2$ using the proposed new algorithm.

R $n_0 = 3, n_1 = 6$			Proposed new algorithm			SPSS	R	RSUE
r	$P(R = r)$	$P(R \leq r)$	Side	One-tailed p -value	Other-tailed p -value	Two-tailed p -value	Two-tailed p -value	Two-tailed p -value
2	0.056	0.056	Left	0.056	0	0.056	0.056	0.111
3	0.194	0.250	Left	0.250	0	0.250	0.250	0.5
4	0.333	0.583	<i>Mdn</i>	0.583	0.417	1	1	1.167
5	0.417	1	Right	0.417	0.417	0.833	0.833	0

Note. r = number of runs, $P(R = r)$ = exact point probability, $P(R \leq r)$ = exact cumulative probability, Side = Left when $r < Mdn(R) = 4$ and Right when $r \geq Mdn(R)$, One-tailed p -value = sum of the exact point probabilities from 2 to r on the left side and from r to the maximum on the right side, Other-tailed p -value = examine the procedure with extreme skewness, since $A_b = -1$ in this distribution, and Two-tailed p -value = sum of the one-tailed and the other-tailed probability values, as when SPSS is used; it is the doubling of one-tailed probability when utilizing R program and RSUE.

The decision can alternatively be based on the critical values or quantiles of the distribution for the Wald-Wolfowitz number of runs with parameters n_0 and n_1 . As the test is two-tailed test, there are two critical values.

${}_{\alpha/2}R_{n_0, n_1}$ = the quantile of order $\alpha/2$ in the Wald-Wolfowitz distribution for the number of runs with parameters n_0 and n_1 , or the critical value in the left tail for a two-tailed test with a significance level α (Equation (18)).

$$P\left(R \leq {}_{\alpha/2}R_{n_0, n_1}\right) \geq \frac{\alpha}{2} \tag{18}$$

${}_{1-\alpha/2}R_{n_0, n_1}$ = the quantile of order $1 - \alpha/2$ in the Wald-Wolfowitz distribution for the number of runs with parameters n_0 and n_1 , or the critical value in the right tail for a two-tailed test with a significance level α (Equation (18)).

$$P\left(R \geq {}_{1-\alpha/2}R_{n_0, n_1}\right) = 1 - P\left(R < {}_{1-\alpha/2}R_{n_0, n_1}\right) = 1 - P\left(R \leq {}_{1-\alpha/2}R_{n_0, n_1} - 1\right) \geq \frac{\alpha}{2} \tag{19}$$

The null hypothesis of randomness is upheld when the number of runs (r) falls within the range of the two critical values, inclusive: ${}_{\alpha/2}R_{n_0, n_1} \leq r \leq {}_{1-\alpha/2}R_{n_0, n_1}$. Conversely, the null hypothesis is rejected if r is below the lower critical value ($r < {}_{\alpha/2}R_{n_0, n_1}$) or exceeds the upper critical value ($r > {}_{1-\alpha/2}R_{n_0, n_1}$). RSUE facilitates the computation of these quantiles. To calculate them using the R program, the following script can be employed [3], where the first number represents the order of the quantile, the second number is the parameter n_0 , the third number is the parameter n_1 , and “lower.tail = TRUE” is used for left-tailed calculations, while “lower.tail = FALSE” is used for right-tailed calculations.

```
library(randtests)
q <- qruns(0.025, 7, 2, lower.tail = TRUE)
cat("0.025_R_7, 3 =", q, "\n")
```

5. Sampling Distribution and Decision with Large-Samples

The distribution of the number of runs tends to converge to the normal distribution [2]. When both n_0 and n_1 exceed 20 and n is greater than or equal to 30, preferably when both n_0 and n_1 exceed 20 and n is greater than 40, the asymptotic probability via the standard normal distribution can be employed [13]. This approximation becomes more accurate when the values of n_0 or n_1 are both close to and large [14].

Initially, the mathematical expectation or arithmetic mean of the number of runs is computed from the normal distribution approximation [1] [2] using the formula in Equation (8). Subsequently, the standard deviation of the number of runs from the normal distribution approximation is determined [1] [2] using the formula in Equation (20).

$$DE(R) = \sqrt{\frac{2 \times n_0 \times n_1 \times (2 \times n_0 \times n_1 - n)}{n^2 \times (n - 1)}} = \sqrt{\frac{(E(R) - 1) \times (E(R) - 2)}{n - 1}} \tag{20}$$

Following that, the test statistic Z , corresponding to the standardized number of runs, is derived [1] [2] using the formula in Equation (21).

$$Z = \frac{r - E(R)}{DE(R)} = \frac{r - \left(\frac{2 \times n_0 \times n_1}{n} + 1\right)}{\sqrt{\frac{2 \times n_0 \times n_1 \times (2 \times n_0 \times n_1 - n)}{n^2 \times (n - 1)}} \tag{21}$$

The Yates continuity correction can be applied by approximating a discrete distribution (number of runs) to a continuous (normal) distribution. This process is automated in the SPSS program [5] through the use of the formula in Equation (22).

$$Z = \frac{(r \mp 0.5) - E(R)}{DE(R)} \tag{22}$$

If $|r - E(R)| \leq 0.5$ or $n \geq 50$, the continuity correction (∓ 0.5) is omitted. For $n < 50$, the continuity correction is applied using the following algorithm: if $r - E(R) < -0.5$, 0.5 is added, and if $r - E(R) > 0.5$, 0.5 is subtracted.

The RSUE program offers an optional continuity correction, while the R program lacks an option for its implementation. It's worth noting that the validity of the continuity correction, particularly the Yates continuity correction, has been a subject of debate [15] [16] [17].

The test statistic Z follows a standard normal distribution: $Z \sim N(0, 1)$. Decisions can rely on the critical values or quantiles of this distribution: if $z_{\alpha/2} \leq z \leq z_{1-(\alpha/2)}$, H_0 holds; on the contrary, if $z < z_{\alpha/2}$ or $z > z_{1-(\alpha/2)}$, H_0 is rejected.

In turn, the decision can be based on the probability value or critical level calculated using the standard normal distribution. When $z \leq 0$, if $2 \times P(Z \leq z) \geq \alpha$, H_0 holds and if $2 \times P(Z \leq z) < \alpha$, H_0 is rejected. When $z > 0$, if $2 \times P(Z \geq z) = 2 \times [1 - P(Z < z)] \geq \alpha$, H_0 holds and if $2 \times [1 - P(Z \leq z)] < \alpha$, H_0 is rejected.

6. Comparing Algorithms for Calculating the Two-Tailed Exact Probability

6.1. Method

Fifty-nine samples with dichotomous data were generated to compare the four algorithms, comprising 29 samples with an even number of data ($n = 10$) and 30 samples with an odd number of data ($n = 11$). See **Table 5**.

On the one hand, nine samples (r from 2 to 10) were derived from a distribution of the number of runs with parameters $n_0 = n_1 = 5$ (R5|5), eight samples (r from 2 to 9) from a distribution with parameters $n_0 = 4$ and $n_1 = 6$ (R4|6), six samples (r from 2 to 7) from a distribution with parameters $n_0 = 3$ and $n_1 = 7$ (R3|7), four samples (r from 2 to 5) from a distribution with parameters $n_0 = 2$ and $n_1 = 8$ (R2|9), and two samples (r from 2 to 3) from a distribution of parameters $n_0 = 1$ and $n_1 = 9$ (R1|9).

On the other hand, ten samples (r from 2 to 11) are drawn from a distribution with parameters $n_0 = 5$ and $n_1 = 6$ (R5|6), eight samples (r from 2 to 9) from a distribution with parameters $n_0 = 4$ and $n_1 = 7$ (R4|7), six samples (r from 2 to 7) from a distribution with parameters $n_0 = 3$ and $n_1 = 8$ (R3|8), four samples (r from 2 to 5) from a distribution with parameters $n_0 = 2$ and $n_1 = 9$ (R2|9), and two samples (r from 2 to 3) from a distribution with parameters $n_0 = 1$ and $n_1 = 10$ (R1|10).

The central tendency comparison of the two-tailed exact probabilities among the four algorithms, whose distributions deviated from normality, was conducted using Friedman's test with a significance level of 5%. Post-hoc comparisons were performed using Conover's test [13] with Benjamini-Yekutielli correction for the familywise error rate [18]. The effect size of the algorithm's impact on the probabilities was estimated by Kendall's W coefficient of concordance and the average Spearman's rank correlation [13].

Table 6 summarizes the computational processes of the four algorithms. It shows the criteria for defining the symmetry axis, how the two sides of the distribution are determined, as well as how the one-tailed and two-tailed probabilities are obtained.

6.2. Results: Calculation and Comparison of Probabilities

Table 7 shows the definition of the (left and right) sides of the 10 distributions of the number of runs. It includes one-tailed and two-tailed probabilities calculated using RSUE and R programs, along with mathematical expectations, exact point, and cumulative probabilities for these distributions. The R program uses mathematical expectations for defining distribution sides, while the RSUE program relies on the median.

Table 8 presents the calculations for one-tailed and two-tailed probabilities using the SPSS programs and the new proposal.

The analysis revealed a significant difference among the four algorithms (Friedman Q -statistic = 58.890, $df = 3$, p -value < 0.001; Iman-Davenport

Table 5. Generated samples.

Sample	Random order										Runs test				
	1	2	3	4	5	6	7	8	9	10	11	n	n_0	n_1	r
R5 5_2r	0	0	0	0	0	1	1	1	1	1		10	5	5	2
R5 5_3r	0	0	0	0	1	1	1	1	1	0		10	5	5	3
R5 5_4r	0	0	1	0	0	0	1	1	1	1		10	5	5	4
R5 5_5r	0	0	1	0	0	1	1	1	1	0		10	5	5	5
R5 5_6r	0	1	0	0	1	1	1	0	0	1		10	5	5	6
R5 5_7r	0	1	0	1	0	0	1	1	1	0		10	5	5	7
R5 5_8r	0	1	0	1	0	0	1	0	1	1		10	5	5	8
R5 5_9r	1	0	1	0	1	0	0	1	0	1		10	5	5	9
R5 5_10r	1	0	1	0	1	0	1	0	1	0		10	5	5	10
R4 6 2r	0	0	0	0	1	1	1	1	1	1		10	4	6	2
R4 6_3r	0	0	0	1	1	1	1	1	1	0		10	4	6	3
R4 6_4r	0	0	1	0	0	1	1	1	1	1		10	4	6	4
R4 6_5r	0	0	1	0	1	1	1	1	1	0		10	4	6	5
R4 6_6r	0	1	0	1	1	1	1	0	0	1		10	4	6	6
R4 6_7r	0	1	0	1	0	1	1	1	1	0		10	4	6	7
R4 6_8r	0	1	0	1	0	1	1	0	1	1		10	4	6	8
R4 6_9r	1	0	1	0	1	0	1	1	0	1		10	4	6	9
R3 7_2r	0	0	0	1	1	1	1	1	1	1		10	3	7	2
R3 7_3r	0	0	1	1	1	1	1	1	1	0		10	3	7	3
R3 7_4r	0	0	1	0	1	1	1	1	1	1		10	3	7	4
R3 7_5r	0	1	1	0	1	1	1	1	1	0		10	3	7	5
R3 7_6r	0	1	0	1	1	1	1	1	0	1		10	3	7	6
R3 7_7r	1	0	1	0	1	0	1	1	1	1		10	3	7	7
R2 8_2r	0	0	1	1	1	1	1	1	1	1		10	2	8	2
R2 8_3r	0	1	1	1	1	1	1	1	1	0		10	2	8	3
R2 8_4r	0	1	1	0	1	1	1	1	1	1		10	2	8	4
R2 8_5r	1	0	1	1	0	1	1	1	1	1		10	2	8	5
R1 9_2r	0	1	1	1	1	1	1	1	1	1		10	1	9	2
R1 9_3r	1	0	1	1	1	1	1	1	1	1		10	1	9	3
R5 6_2r	0	0	0	0	0	1	1	1	1	1	1	11	5	6	2

Continued

R5 6_3r	0	0	0	0	1	1	1	1	1	1	0	11	5	6	3
R5 6_4r	0	0	0	1	0	0	1	1	1	1	1	11	5	6	4
R5 6_5r	0	0	0	1	0	1	1	1	1	1	0	11	5	6	5
R5 6_6r	0	0	1	0	1	1	1	1	0	0	1	11	5	6	6
R5 6_7r	0	0	1	0	1	0	1	1	1	1	0	11	5	6	7
R5 6_8r	0	0	1	0	1	0	1	1	0	1	1	11	5	6	8
R5 6_9r	1	0	0	1	0	1	0	1	0	1	1	11	5	6	9
R5 6_10r	1	0	1	0	1	0	1	1	0	1	0	11	5	6	10
R5 6_11r	1	0	1	0	1	0	1	0	1	0	1	11	5	6	11
R4 7_2r	0	0	0	0	1	1	1	1	1	1	1	11	4	7	2
R4 7_3r	0	0	0	1	1	1	1	1	1	1	0	11	4	7	3
R4 7_4r	0	0	1	0	0	1	1	1	1	1	1	11	4	7	4
R4 7_5r	0	0	1	1	0	1	1	1	1	1	0	11	4	7	5
R4 7_6r	0	1	1	0	1	1	1	1	0	0	1	11	4	7	6
R4 7_7r	0	1	1	0	1	0	1	1	1	0	1	11	4	7	7
R4 7_8r	0	1	1	0	1	0	1	1	0	1	1	11	4	7	8
R4 7_9r	1	0	1	1	0	1	0	1	1	0	1	11	4	7	9
R3 8_2r	0	0	0	1	1	1	1	1	1	1	1	11	3	8	2
R3 8_3r	0	0	1	1	1	1	1	1	1	1	0	11	3	8	3
R3 8_4r	0	1	1	1	0	0	1	1	1	1	1	11	3	8	4
R3 8_5r	0	1	1	1	0	1	1	1	1	1	0	11	3	8	5
R3 8_6r	0	1	1	0	1	1	1	1	1	0	1	11	3	8	6
R3 8_7r	1	0	1	1	0	1	0	1	1	1	1	11	3	8	7
R2 9_2r	0	0	1	1	1	1	1	1	1	1	1	11	2	9	2
R2 9_3r	0	1	1	1	1	1	1	1	1	1	0	11	2	9	3
R2 9_4r	0	1	1	1	1	0	1	1	1	1	1	11	2	9	4
R2 9_5r	1	0	1	1	1	0	1	1	1	1	1	11	2	9	5
R1 10_2r	0	0	1	1	1	1	1	1	1	1	1	11	1	10	2
R1 10_3r	1	0	1	1	1	1	1	1	1	1	1	11	1	10	3

Note. Each sample is identified by the parameters n_0 and n_1 of the R distribution from which it is drawn and its number of runs (r). Runs test: $n = n_0 + n_1 =$ sample size, $n_0 =$ number of zeros, $n_1 =$ number of ones, $r =$ number of runs (threshold: 0.5).

Table 6. Synthesis of the calculation of the one- or two-tailed exact probability using the four algorithms.

Software	Axis of symmetry	Side R n_0, n_1	$P_{1 \text{ tail}}$	$P_{2 \text{ tails}}$
SPSS	$E(R)$	left: $r \leq E(R)$ right: $r > E(R)$	$P(2 \leq R \leq r)$ $P(r \leq R \leq \text{Max}(R))$	$r_{\text{other side}} \geq r - E(R) $ $P_{\text{other tail}} = P_{1 \text{ tail}}(r_{\text{other side}})$ $P_{2 \text{ tails}} = P_{1 \text{ tail}} + P_{\text{other tail}}$
R	$E(R)$	left: $r \leq E(R)$ right: $r > E(R)$	$P(2 \leq R \leq r)$ $P(r \leq R \leq \text{Max}(R))$	$P_{2 \text{ tails}} = 2 \times P_{1 \text{ tail}}$
RSUE	$Mdn(R)$	left: $r < Mdn(R)$ right: $r \geq Mdn(R)$	$P(2 \leq R \leq r)$ $P(r \leq R \leq \text{Max}(R))$	$P_{2 \text{ tails}} = 2 \times P_{1 \text{ tail}}$ $P_{2 \text{ tails}} = 2 \times (1 - P_{1 \text{ tail}})$
Proposal	$Mdn(R)$	left: $r < Mdn(R)$ right: $r \geq Mdn(R)$ $r = Mdn(R)$	$P(2 \leq R \leq r)$ $P(r \leq R \leq \text{Max}(R))$ $P(R \leq Mdn(R))$	$P_{\text{other tail}} \geq P_{1 \text{ tail}}^*$ $P_{2 \text{ tails}} = P_{1 \text{ tail}} + P_{\text{other tail}}$ $P_{\text{other tail}} = 1 - P(R \leq Mdn(R))$ $P_{2 \text{ tails}} = P_{1 \text{ tail}} + P_{\text{other tail}} = 1$

Note. RSUE = Real Statistics using Excel, r = number of runs, varying from 2 to $\text{Max}(R) = n_0 + n_1$ when $n_0 = n_1$ or $2 \times \min(n_0, n_1) + 1$ when $n_0 \neq n_1$, $E(R)$ = mathematical expectation and $Mdn(R)$ = median of the number of runs distribution with parameters n_0 and n_1 .
* If there is no equal or greater probability, the one-tailed probability is doubled to obtain two-tailed probability, unless there is extreme negative skewness ($A_B = -1$). In this case, the probabilities corresponding to the other tail of the values of R in the lengthened tail are 0 and, with the value of R in the shortened tail, the one-tailed probability is doubled with the restriction that the maximum value of the two-tailed probability is 1.

Table 7. One-tailed and two-tailed exact probability with RSUE and R programs.

Sample	Mean	Exact probability			RSUE		R program		
		Point	Cum.	Side	1 tail	2 tails	Side	1 tail	2 tails
R5 5_2r	6	0.008	0.008	left	0.008	0.016	left	0.008	0.016
R5 5_3r	6	0.032	0.040	left	0.040	0.079	left	0.040	0.079
R5 5_4r	6	0.127	0.167	left	0.167	0.333	left	0.167	0.333
R5 5_5r	6	0.190	0.357	left	0.357	0.714	left	0.357	0.714
R5 5_6r	6	0.286	0.643	right	0.357	0.714	left	0.643	1.286
R5 5_7r	6	0.190	0.833	right	0.167	0.333	right	0.357	0.714
R5 5_8r	6	0.127	0.960	right	0.040	0.079	right	0.167	0.333
R5 5_9r	6	0.032	0.992	right	0.008	0.016	right	0.040	0.079
R5 5_10r	6	0.008	1	right	0	0	right	0.008	0.016
R4 6 2r	5.8	0.010	0.010	left	0.010	0.019	left	0.010	0.019
R4 6_3r	5.8	0.038	0.048	left	0.048	0.095	left	0.048	0.095
R4 6_4r	5.8	0.143	0.190	left	0.190	0.381	left	0.191	0.381

Continued

R4 6_5r	5.8	0.214	0.405	left	0.405	0.810	left	0.405	0.810
R4 6_6r	5.8	0.286	0.690	right	0.310	0.619	right	0.595	1.190
R4 6_7r	5.8	0.190	0.881	right	0.119	0.238	right	0.310	0.619
R4 6_8r	5.8	0.095	0.976	right	0.024	0.048	right	0.119	0.238
R4 6_9r	5.8	0.024	1	right	0	0	right	0.024	0.048
R3 7_2r	5.2	0.017	0.017	left	0.017	0.033	left	0.017	0.033
R3 7_3r	5.2	0.067	0.083	left	0.083	0.167	left	0.083	0.167
R3 7_4r	5.2	0.2	0.283	left	0.283	0.567	left	0.283	0.567
R3 7_5r	5.2	0.3	0.583	right	0.417	0.833	left	0.583	1.167
R3 7_6r	5.2	0.25	0.833	right	0.167	0.333	right	0.417	0.833
R3 7_7r	5.2	0.167	1	right	0	0	right	0.167	0.333
R2 8_2r	4.2	0.044	0.044	left	0.044	0.089	left	0.044	0.089
R2 8_3r	4.2	0.178	0.222	left	0.222	0.444	left	0.222	0.444
R2 8_4r	4.2	0.311	0.533	left	0.467	0.933	left	0.533	1.067
R2 8_5r	4.2	0.467	1	right	0	0	right	0.467	0.933
R1 9_2r	2.8	0.2	0.200	left	0.2	0.4	left	0.2	0.4
R1 9_3r	2.8	0.8	1	right	0	0	right	0.8	1.6
R5 6_2r	6.455	0.004	0.004	left	0.004	0.009	left	0.004	0.009
R5 6_3r	6.455	0.019	0.024	left	0.024	0.048	left	0.024	0.048
R5 6_4r	6.455	0.087	0.110	left	0.110	0.221	left	0.110	0.221
R5 6_5r	6.455	0.152	0.262	left	0.262	0.524	left	0.262	0.524
R5 6_6r	6.455	0.260	0.522	right	0.478	0.957	left	0.522	1.043
R5 6_7r	6.455	0.216	0.738	right	0.262	0.524	right	0.478	0.957
R5 6_8r	6.455	0.173	0.911	right	0.089	0.177	right	0.262	0.524
R5 6_9r	6.455	0.065	0.976	right	0.024	0.048	right	0.089	0.177
R5 6_10r	6.455	0.022	0.998	right	0.002	0.004	right	0.024	0.048
R5 6_11r	6.455	0.002	1	right	0	0	right	0.002	0.004
R4 7_2r	6.091	0.006	0.006	left	0.006	0.012	left	0.006	0.012
R4 7_3r	6.091	0.027	0.033	left	0.033	0.067	left	0.033	0.067
R4 7_4r	6.091	0.109	0.142	left	0.142	0.285	left	0.142	0.285
R4 7_5r	6.091	0.191	0.333	left	0.333	0.667	left	0.333	0.667
R4 7_6r	6.091	0.273	0.606	right	0.394	0.788	left	0.606	1.212
R4 7_7r	6.091	0.227	0.833	right	0.045	0.091	right	0.167	0.333
R4 7_8r	6.091	0.121	0.955	right	0.045	0.091	right	0.167	0.333

Continued

R4 7_9r	6.091	0.045	1.000	right	0	0	right	0.045	0.091
R3 8_2r	5.364	0.012	0.012	left	0.012	0.024	left	0.012	0.024
R3 8_3r	5.364	0.055	0.067	left	0.067	0.133	left	0.067	0.133
R3 8_4r	5.364	0.170	0.236	left	0.236	0.473	left	0.236	0.473
R3 8_5r	5.364	0.297	0.533	right	0.467	0.933	left	0.533	1.067
R3 8_6r	5.364	0.255	0.788	right	0.212	0.424	right	0.467	0.933
R3 8_7r	5.364	0.212	1	right	0	0	right	0.212	0.424
R2 9_2r	4.273	0.036	0.036	left	0.036	0.073	left	0.036	0.073
R2 9_3r	4.273	0.164	0.200	left	0.200	0.400	left	0.2	0.4
R2 9_4r	4.273	0.291	0.491	left	0.491	0.982	left	0.491	0.982
R2 9_5r	4.273	0.509	1	right	0	0	right	0.509	1.018
R1 10_2r	2.818	0.182	0.182	left	0.036	0.073	left	0.036	0.073
R1 10_3r	2.818	0.818	1	right	0	0	right	0.818	1.636

Note. Each sample is identified by the parameters n_0 and n_1 of the R distribution from which it is drawn and its number of runs (r), $n = n_0 + n_1 =$ sample size, $n_0 =$ number of zeros, $n_1 =$ number of ones, $r =$ number of runs (threshold: 0.5).

Table 8. One-tailed and two-tailed exact probability with RSUE and R programs.

Sample	SPSS						New proposal			
	1	$ r - E(R) $	Other	Other	2	1	Sk_B	Side	Other	2
	tail		side	side	tails	tails			side	tails
P	r	P	P	P	P	P	P	P	P	
R5 5_2r	0.008	4	10	0.008	0.016	0.008	-0.286	left	0.008	0.016
R5 5_3r	0.040	3	9	0.040	0.079	0.040		left	0.040	0.079
R5 5_4r	0.167	2	8	0.167	0.333	0.167		left	0.167	0.333
R5 5_5r	0.357	1	7	0.357	0.714	0.357		left	0.357	0.714
R5 5_6r	0.643	0	7	0.357	1	0.643		<i>Mdn</i>	0.357	1
R5 5_7r	0.357	1	5	0.357	0.714	0.357		right	0.357	0.714
R5 5_8r	0.167	2	4	0.167	0.333	0.167		right	0.167	0.333
R5 5_9r	0.040	3	3	0.040	0.079	0.040		right	0.040	0.079
R5 5_10r	0.008	4	2	0.008	0.016	0.008		right	0.008	0.016
R4 6 2r	0.010	3.8	0	0	0.010	0.010	-0.381	left	0.024	0.033
R4 6_3r	0.048	2.8	9	0.024	0.071	0.048		left	0.119	0.167
R4 6_4r	0.190	1.8	8	0.119	0.310	0.190		left	0.310	0.5
R4 6_5r	0.405	0.8	7	0.310	0.714	0.405		left	0.405	0.810
R4 6_6r	0.690	0.2	7	0.310	1	0.690		<i>Mdn</i>	0.310	1

Continued

R4 6_7r	0.310	1.2	4	0.190	0.5	0.310		right	0.405	0.714
R4 6_8r	0.119	2.2	3	0.048	0.167	0.119		right	0.190	0.310
R4 6_9r	0.024	3.2	2	0.010	0.033	0.024		right	0.048	0.071
R3 7_2r	0.017	3.2	0	0	0.017	0.017	-0.167	left	0.167	0.183
R3 7_3r	0.083	2.2	0	0	0.083	0.083		left	0.167	0.25
R3 7_4r	0.283	1.2	7	0.1667	0.45	0.283		left	0.417	0.7
R3 7_5r	0.583	0.2	6	0.4167	1	0.583		<i>Mdn</i>	0.417	1
R3 7_6r	0.417	0.8	4	0.2833	0.7	0.417		right	0.417	0.833
R3 7_7r	0.167	1.8	3	0.0833	0.25	0.167		right	0.283	0.45
R2 8_2r	0.044	2.2	0	0	0.044	0.044	-1	left	0	0.044
R2 8_3r	0.222	1.2	0	0	0.222	0.222		left	0	0.222
R2 8_4r	0.533	0.2	5	0.467	1	0.533		<i>Mdn</i>	0.467	1
R2 8_5r	0.467	0.8	3	0.222	0.689	0.467		right	0.467	0.933
R1 9_2r	0.2	0.8	0	0	0.2	0.2	-1	left	0	0.2
R1 9_3r	0.8	0.2	2	0.2	1	0.8		<i>Mdn</i>	0.2	1
R5 6_2r	0.004	4.455	11	0.002	0.006	0.004	-0.043	left	0.024	0.028
R5 6_3r	0.024	3.455	10	0.024	0.048	0.024		left	0.024	0.048
R5 6_4r	0.110	2.455	9	0.089	0.199	0.110		left	0.262	0.372
R5 6_5r	0.262	1.455	8	0.262	0.524	0.262		left	0.262	0.524
R5 6_6r	0.522	0.455	7	0.478	1	0.522		<i>Mdn</i>	0.478	1
R5 6_7r	0.478	0.545	5	0.262	0.740	0.478		right	0.478	0.957
R5 6_8r	0.262	1.545	4	0.110	0.372	0.262		right	0.262	0.524
R5 6_9r	0.089	2.545	3	0.024	0.113	0.089		right	0.110	0.199
R5 6_10r	0.024	3.545	2	0.004	0.028	0.024		right	0.024	0.048
R5 6_11r	0.002	4.545	0	0	0.002	0.002		right	0.004	0.006
R4 7_2r	0.006	4.091	0	0	0.006	0.006	-0.212	left	0.045	0.052
R4 7_3r	0.033	3.091	0	0	0.033	0.033		left	0.045	0.079
R4 7_4r	0.142	2.091	9	0.045	0.188	0.142		left	0.167	0.309
R4 7_5r	0.333	1.091	8	0.167	0.5	0.333		left	0.394	0.727
R4 7_6r	0.606	0.091	7	0.394	1	0.606		<i>Mdn</i>	0.394	1
R4 7_7r	0.394	0.909	5	0.333	0.727	0.394		right	0.394	0.788
R4 7_8r	0.167	1.909	4	0.142	0.309	0.167		right	0.333	0.5
R4 7_9r	0.045	2.909	3	0.033	0.079	0.045		right	0.142	0.188
R3 8_2r	0.012	3.364	0	0	0.012	0.012	-0.067	left	0.212	0.224
R3 8_3r	0.067	2.364	0	0	0.067	0.067		left	0.212	0.279

Continued

R3 8_4r	0.236	1.364	7	0.212	0.448	0.236		left	0.467	0.703
R3 8_5r	0.533	0.364	6	0.467	1	0.533		<i>Mdn</i>	0.467	1
R3 8_6r	0.467	0.636	4	0.236	0.703	0.467		right	0.467	0.933
R3 8_7r	0.212	1.636	3	0.067	0.279	0.212		right	0.236	0.448
R2 9_2r	0.036	2.273	0	0	0.036	0.036	-1	left	0	0.036
R2 9_3r	0.2	1.273	0	0	0.2	0.2		left	0	0.2
R2 9_4r	0.491	0.273	5	0.509	1	0.491		left	0	0.491
R2 9_5r	0.509	0.727	3	0.2	0.709	0.509		<i>Mdn</i>	0.491	1
R1 10_2r	0.182	0.818	0	0	0.182	0.182	-1	left	0	0.182
R1 10_3r	0.818	0.182	2	0.182	1	0.818		<i>Mdn</i>	0.182	1

Note. Sample: $n_0|n_1-r$ = parameters: n_0 (numbers of 0 s) and n_1 (numbers of 1 s), and r = number of runs, p = probability value. $|r - E(R|n_0, n_1)|$ = absolute difference of the number of runs r and the mathematical expectation of the distribution of the number of runs; a horizontal line separates the two sides of the distribution. *Mdn* ($R|R|n_0, n_1$) = median of the distribution of the number of runs. Sk_B = Bickel's skewness coefficient based on the mode.

Table 9. Differences between the two-tailed exact probabilities of the four procedures.

Statistics	E - R	E - SPSS	E - N	R - SPSS	R - N	SPSS - N	E - R
	<0	31	33	48	3	27	34
Frequency	=0	28	6	7	10	16	26
	>0	2	22	6	48	18	1
	<i>MD</i>	-0.205	-0.118	-0.199	0.087	0.006	-0.081
	<i>DMR</i>	1.212	0.127	1.373	1.085	0.161	1.246
Post-hoc comparison using Conover test	<i>SE</i>	0.188	0.188	0.188	0.188	0.188	0.188
	<i>t</i> -statistic	6.451	0.677	7.308	5.774	0.857	6.631
	<i>p</i> -value	0.000	0.500	0.000	0.000	0.393	0.000
	α_c	0.014	0.003	0.020	0.010	0.007	0.017
	Significance	yes	no	yes	yes	no	yes
	Direction	E < R	E = SPSS	E < N	SPSS < R	R = N	SPSS < N

Note. E = Real Statistics using Excel, R = R program, SPSS = Statistical Package for the Social Sciences, N = new proposal. Frequency: <0 (number of negative differences), = 0 (number of null differences or equivalences), and >0 (number of positive differences). Pairwise comparisons using Friedman-Conover test with Benjamini-Yekutielli correction for familywise error: *MD* = mean signed differences of two-tailed exact probabilities, *DRM* = difference in mean ranks, *SE* = standard error of difference in mean ranks, *t*-statistic = $DRM/SE = DRM/0.188$, *p*-value = two-tailed probability value following a t-distribution with 174 degree of freedom, α_c = corrected significance level, significance: "yes" when *p*-value is less than α_c and "no" when *p*-value is greater or equal than α_c .

F-statistic = 28.919, $df_1 = 3$, $df_2 = 174$, p -value < 0.001). The effect size was medium (Kendall's $w = 0.333$, $\bar{r}_s = 0.321$). Utilizing Conover's test with the Benjamini-Yekutieli correction, significant differences of central tendency in probability were observed between RSUE and the R program (RSUE < R), as well as the new proposal (RSUE < NP), and similarly between SPSS and the R program (SPSS < R) and the new proposal (SPSS < NP). The central tendencies in probability between RSUE and SPSS, as well as between R program and the new proposal, were found to be statistically equivalent (Table 9).

7. Conclusions

It is important to acknowledge that the computation employed by RSUE assumes symmetry, a condition valid only when $n_1 = n_2$. Moreover, RSUE treats the distribution of the number of runs as an analog of the normal distribution, which is continuous and symmetrical, when this distribution is discrete and exceptionally symmetrical. Accordingly, RSUE uses the median as axis of symmetry instead of the mathematical expectation. This analogy may be based on the distributional convergence to the normal distribution [2]. However, with small samples, the approximation is not appropriate [14]. When employing this algorithm, it emerges that both the one-tailed and two-tailed probabilities of the maximum value of R are zero. This implies that the lowest mean results when averaging the exact two-tailed probabilities of the 59 samples. It should be noted that the fact that the one-tailed and two-tailed probability is always zero at the maximum r value when the point probability is not zero is a contradiction, and constitutes a further inconsistency with this program.

The SPSS program's algorithm employs the mathematical expectation as the axis of symmetry to delineate the two sides of the distribution and determine the value of r corresponding to the opposite side. The sum of the one-tailed probabilities of r and its counterpart on the other side yields the two-tailed probability. The proposed algorithm shares similarities but defines the distribution's two sides using the median. It employs quantile order on the left side and the complementary quantile order on the right to ascertain the probability corresponding to the opposite tail, except for the median, where its complement represents the probability of the other tail. The summation of both probabilities provides the exact two-tailed probability, which is more accurate given the varied asymmetry situations presented by the distribution of the number of runs, including extreme negative asymmetry [9]. However, both procedures are more suitable than doubling the one-sided probability, as performed by the R program. Doubling the one-sided probability fails to adequately consider the asymmetry characteristics of the distribution of the number of runs and may result in probability values exceeding 1.

In the central tendency comparisons, the proposed algorithm is slightly more conservative with the null hypothesis of randomness than the SPSS program. It is a general characteristic of robust procedures [19], although it may be per-

ceived as a drawback in terms of decisions relying on exact probability [20] [21] [22].

This study generated various samples, utilizing a predefined cut-off point to ensure that the sum of n_0 and n_1 consistently equaled the sample size (n) across different programs. The goal was to simulate the behavior of the distribution of the number of runs. Thus, when $n_0 = n_1$, the exact two-tailed probabilities of the four procedures coincide, except in the case of the RSUE when $r = n = 10$.

However, even when $n_0 = n_1$, there are discrepancies in the exact two-tailed probabilities between RSUE, SPSS, and R program. This discrepancy arises because R program and RSUE exclude sample values equal to the cut-off point, whereas SPSS does not. The exclusion of values equal to the cut-off point is justified as a correction for ties, a common practice when assigning ranks [23]. This correction aims to enhance the power of non-parametric tests, akin to the continuity correction. However, its effectiveness remains debatable [24], leading to varied implementation across programs, with some adopting it and others abstaining [4].

An option not explored in this study is the calculation of probability through bootstrapping. The SPSS program includes this functionality, albeit limited to the Z statistic of the asymptotic approximation, which is suitable for large samples. For small samples, the true choice is exact probability [3] [6], as bootstrap is not recommended for small samples [25]. This underscores the significance of the present study.

As limitations of the study, it should be noted that only three programs have been considered, though others like C#, BMDP, SAS, STATISTX, GNU/Octave also include the runs test. Moreover, two sample sizes were chosen: even (10 data points) and odd (11 data points). This simplification is assumed to be adequate, representative, and clear for comparing exact two-tailed probabilities computed by the four algorithms.

In conclusion, the calculation provided by SPSS and the new proposal is deemed more suitable when compared to the calculations of the RSUE and R programs. The new proposal represents a robust and well-supported calculation procedure, albeit slightly more conservative than the SPSS method. Additionally, it has the potential for generalization to other tests, including binomial and sign tests.

Acknowledgements

The author thanks the reviewers and editor for their helpful comments.

Conflicts of Interest

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

References

- [1] Wald, A. and Wolfowitz, J. (1943) An Exact Test for Randomness in the Non-Para-

- metric Case Based on Serial Correlation. *The Annals of Mathematical Statistics*, **14**, 378-388. <https://doi.org/10.1214/aoms/1177731358>
- [2] Wald, A. and Wolfowitz, J. (1940) On a Test Whether Two Samples Are from the Same Population. *The Annals of Mathematical Statistics*, **11**, 147-162. <https://doi.org/10.1214/aoms/1177731909>
- [3] Caeiro, F. (2022) Package “Randtests”. <https://cran.r-project.org/web/packages/randtests/randtests.pdf>
- [4] Zaiantz, C. (2016) One-Sample Runs Test. Real Statistics Using Excel. <https://real-statistics.com/non-parametric-tests/one-sample-runs-test/>
- [5] IBM Corporation (2013) IBM SPSS Statistics 22 Algorithms. https://www.sussex.ac.uk/its/pdfs/SPSS_Statistics_Algorithms_22.pdf
- [6] IBM Corporation (2022) SPSS Statistics 29. Documentation. <https://www.ibm.com/docs/en/spss-statistics/29.0.0>
- [7] Mehta, C.R. and Patel, N.R. (2013) IBM SPSS Exact Tests. https://www.ibm.com/docs/en/SSLVMB_27.0.0/pdf/en/IBM_SPSS_Exact_Tests.pdf
- [8] Mohajan, H.K. (2020) Quantitative Research: A Successful Investigation in Natural and Social Sciences. *Journal of Economic Development, Environment and People*, **9**, 50-79.
- [9] Gunver, M.G., Senocak, M.S. and Vehid S. (2018) To Determine Skewness, Mean and Deviation with a New Approach on Continuous Data. *Ponte: International Scientific Researches Journal*, **73**, 64-79. <https://doi.org/10.21506/j.ponte.2018.2.5>
- [10] Verma, J.P. and Abdel-Salam, A.-S.G. (2019) Testing Statistical Assumptions in Research. John Wiley and Sons, Hoboken.
- [11] Poncet, P. (2022) Package Modeest. Mode Estimation. <https://cran.r-project.org/web/packages/modeest/modeest.pdf>
- [12] Bickel, D.R. (2002) Robust Estimators of the Mode and Skewness of Continuous Data. *Computational Statistics & Data Analysis*, **39**, 153-163. [https://doi.org/10.1016/S0167-9473\(01\)00057-3](https://doi.org/10.1016/S0167-9473(01)00057-3)
- [13] Conover, W.J. (1999) Practical Nonparametric Statistics. 3th Edition, John Wiley and Sons, Hoboken.
- [14] Diamantopoulos, A., Schlegelmilch, B. and Halkias, G. (2023) Simple Things First: One Variable, One Sample. In: *Taking the Fear out of Data Analysis*, Edward Elgar Publishing, Cheltenham, 163-184. <https://doi.org/10.4337/9781803929842.00022>
- [15] Alvo, M. (2022) Nonparametric Statistics. In: *Statistical Inference and Machine Learning for Big Data*, Springer International Publishing, Cham, 95-170.
- [16] Kvam, P., Vidakovic, B. and Ki, S.-J. (2022) Nonparametric Statistics with Applications to Science and Engineering with R. John Wiley and Sons, Hoboken.
- [17] Martín Andrés, A., Álvarez Hernández, M. and Gayá Moreno, F. (2023) The Yates, Conover, and Mantel Statistics in 2×2 Tables Revisited (and Extended). *Statistica Neerlandica*. <https://doi.org/10.1111/stan.12320>
- [18] Benjamini, Y. and Yekutieli, D. (2001) The Control of False Discovery Rate in Multiple Testing under Dependency. *Annals of Statistics*, **29**, 1165-1188. <https://doi.org/10.1214/aos/1013699998>
- [19] Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M. (2019) Robust Statistics: Theory and Methods (with R). 2nd Edition, John Wiley & Sons, Hoboken.
- [20] Bind, M.A. and Rubin, D.B. (2020) When Possible, Report a Fisher-Exact p Value and Display Its Underlying Null Randomization Distribution. *Proceedings of the*

National Academy of Sciences, **117**, 19151-19158.

<https://doi.org/10.1073/pnas.1915454117>

- [21] DiCiccio, C.J., DiCiccio, T.J. and Romano, J.P. (2020) Exact Tests via Multiple Data Splitting. *Statistics & Probability Letters*, **166**, Article ID: 108865.
<https://doi.org/10.1016/j.spl.2020.108865>
- [22] Frey, J. and Zhang, Y. (2019) Improved Exact Confidence Intervals for a Proportion Using Ranked-Set Sampling. *Journal of the Korean Statistical Society*, **48**, 493-501.
<https://doi.org/10.1016/j.jkss.2019.05.003>
- [23] Hodges, C.B., Stone, B.M., Johnson, P.K., Carter III, J.H., Sawyers, C.K., Roby, P.R. and Lindsey, H.M. (2023) Researcher Degrees of Freedom in Statistical Software Contribute to Unreliable Results: A Comparison of Nonparametric Analyses Conducted in SPSS, SAS, Stata, and R. *Behavior Research Methods*, **55**, 2813-2837.
<https://doi.org/10.3758/s13428-022-01932-2>
- [24] McGee, M. (2018) Case for Omitting Tied Observations in the Two-Sample T-Test and the Wilcoxon-Mann-Whitney Test. *PLOS ONE*, **13**, e0200837.
<https://doi.org/10.1371/journal.pone.0200837>
- [25] Rousselet, G., Pernet, C.R. and Wilcox, R.R. (2023) An Introduction to the Bootstrap: A Versatile Method to Make Inferences by Using Data-Driven Simulations. *Meta-Psychology*, **7**, Article No. 2058.
<https://doi.org/10.15626/MP.2019.2058>