Scientific Research Publishing

Open Journal of Statistics, 2024, 14, 119-149
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Rice University Rule to Determine the Number of Bins

José Moral De La Rubia

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

How to cite this paper: Rubia, J.M.D.L. (2024) Rice University Rule to Determine the Number of Bins. *Open Journal of Statistics*, **14**, 119-149.
https://doi.org/10.4236/ojs.2024.141006

Received: December 16, 2023
Accepted: February 26, 2024
Published: February 29, 2024

Abstract

This study aims to establish a rationale for the Rice University rule in determining the number of bins in a histogram. It is grounded in the Scott and Freedman-Diaconis rules. Additionally, the accuracy of the empirical histogram in reproducing the shape of the distribution is assessed with respect to three factors: the rule for determining the number of bins (square root, Sturges, Doane, Scott, Freedman-Diaconis, and Rice University), sample size, and distribution type. Three measures are utilized: the average distance between empirical and theoretical histograms, the level of recognition by an expert judge, and the accuracy index, which is composed of the two aforementioned measures. Mean comparisons are conducted with aligned rank transformation analysis of variance for three fixed-effects factors: sample size (20, 35, 50, 100, 200, 500, and 1000), distribution type (10 types), and empirical rule to determine the number of bins (6 rules). From the accuracy index, Rice's rule improves with increasing sample size and is independent of distribution type. It outperforms the Friedman-Diaconis rule but falls short of Scott's rule, except with the arcsine distribution. Its profile of means resembles the square root rule concerning distributions and Doane's rule concerning sample sizes. These profiles differ from those of the Scott and Friedman-Diaconis rules, which resemble each other. Among the seven rules, Scott's rule stands out in terms of accuracy, except for the arcsine distribution, and the square root rule is the least accurate.

Keywords

Histogram, Class Intervals, Accuracy, Distributions, Descriptive Statistics

1. Introduction

Lane [1] presents the Rice University rule as a straightforward and practical me-

thod that appears to perform well across various sample sizes without assuming any specific distribution. Additionally, the author asserts that it enables the effective visualization of distribution shapes [2]. In practice, the rule is employed in applied research [3]. However, the rationale behind the rule is not provided, and it has received limited attention in the literature. One of the few studies investigating this rule is conducted by Sahann, Müller, and Schmidt [4]. In their research, they explore the correlation between the number of intervals required for a user to recognize a distribution and the number of intervals generated by four empirical rules. They conclude that the Scott, Rice University, and Freedman-Diaconis rules tend to overestimate the number of intervals. The Sturges rule emerges as the most suitable among the four, and they ultimately recommend using a fixed number of 20 class intervals. Their study incorporates four different distributions: uniform, normal, bimodal, and gamma, and four sample sizes: 100, 1000, 10,000, and 1,000,000. The user sample comprises 100 computer science students. Consequently, in this validation study, Rice's rule does not stand out as a preferable choice.

The present study aims to provide a rationale for the Rice University rule based on the Scott and Freedman-Diaconis rules. Additionally, it seeks to assess the accuracy of the empirical histogram regarding six rules for determining the number of bins, including five classical rules: Pearson's square root [5], Sturges [6], Doane [7], Scott [8], and Freedman-Diaconis [9], along with the Rice University rule [1]. The study does not initially emphasize one rule over another. The assessment includes two additional factors: sample size and distribution type, considering both their principal effects and the interactions of second and third order among the three factors.

It begins with the presentation of a historical note on the histogram and six empirical rules to determine the number of class intervals. When the Rice University rule [1] is presented, its rationale is developed from the rules of Scott [8] and Freedman-Diaconis [9].

The second objective of the study is pursued by randomly drawing five distinct samples from a standard continuous uniform distribution across seven different sizes: 20, 35, 50, 100, 200, 500, and 1000 data points. Employing the inverse transform sampling method, a total of 350 samples are generated, incorporating 10 distribution types. These distribution types encompass two mesokurtic symmetric distributions: normal N ($\mu = 5$, $\sigma^2 = 6.25$) and beta ($\alpha = 30$, $\beta = 30$); two leptokurtic symmetric distributions: Laplace ($\mu = 5$, $\beta = 2.5$) and logistic ($\mu = 5$, $s = 2.5$); two platykurtic symmetric distributions: arcsine ($a = 1$, $b = 1$) and semicircular ($R = 2$); two distributions exhibiting skewness and platykurtosis: triangular ($a = b = 0$, $c = 1$) and PERT ($a = 1$, $b = 4$, $c = 5$); and two distributions featuring skewness and leptokurtosis: exponential ($\lambda = 1/2$) and lognormal ($\mu = 0$, $\sigma^2 = 0.25$). Subsequently, the six rules are applied to these 350 samples, resulting in the generation of 2100 empirical histograms (5 samples of 7 sizes of 10 distributions for 6 rules).

The probability of k class intervals in each of the 2100 empirical histograms was calculated using the cumulative distribution function of the corresponding distribution. This process generated the expected relative frequencies or theoretical probabilities. The distance between the relative frequencies and the expected probabilities of the class intervals was then measured using the average Euclidean distance, referred to as the *average discrepancy*.

On the other hand, a sample of 1000 uniformly distributed data with perfect symmetry was generated in the interval [0.001, 0.999], equispaced at a distance of 0.000998. Using the corresponding quantile function, 10 distributions were generated. A probability density function plot for each of the 10 distributions, bounded to the range of the random sample, served as a theoretical model to visually assess whether the empirical histogram accurately reproduces the distribution curve. The visual evaluation was conducted by an expert judge using a scale of four ordered categories, referred to as *recognition level*.

The means of the average discrepancy and the recognition level are compared, taking into account three factors: rule, distribution, and sample size, through analysis of variance with aligned rank transformation. Additional details are provided in the Method section. Following the explanation of the study's methodology, the Result section is presented, and conclusions are subsequently drawn.

2. Historical Note on the Histogram

In 1833, the histogram was introduced by the French lawyer and statistician André Michel Guerry [10] as an approximation of a discrete empirical distribution to a continuous distribution function in the study of crimes and suicides in France. The English nurse Florence Nightingale [11] utilized histograms to compare the mortality of soldiers and civilians in her work on sanitation in the British army during the Crimean War against Russia, published in 1859. However, these authors simply named their graphic representations and did not use the neologism "histogram", composed of the Greek words "histos" ($\iota\sigma\tau o\sigma$) (which can be translated into English as "mast") and "gramma" ($\gamma\rho\alpha\mu\mu\alpha$) (which can be translated as "graph" or "drawing"). Thus, this term, in its etymological sense, refers to a graph of masts or vertical bars [12].

The statistical term "histogram" was coined by Karl Pearson (1857-1936) and first used in his lecture on maps and cartograms in 1891, during his tenure as a professor of geometry at Gresham College. In this lecture, Pearson explained that histograms could be used to represent historical information about reigns, sovereigns, or prime ministers of different periods [13]. However, it wasn't until 1895 that the term "histogram" appeared in a written publication by Pearson when he presented his system of continuous distributions. In contrast to Guerry, Pearson discretizes a continuum of values into k class intervals to create the attached rectangles seen in the histogram, representing the areas under the continuous distribution [14].

It is noteworthy that the histogram serves as a highly valuable graphical tool for comprehending the shape of a distribution and found extensive use in Karl Pearson's development of his system of frequency curves [14]. The histogram allows for the graphical representation of the sampling distribution of a continuous random variable. This involves grouping the n sample values within a continuous range into k class intervals. Each interval is represented in the histogram by a rectangle, with its area corresponding to its relative frequency.

On the horizontal or abscissa axis, the k intervals are arranged and can be labeled with their two limits in brackets or parentheses $[LI, LS)$ or with a singular value (the class mark). This axis can take values from 0 to $+\infty$, from $-\infty$ to $+\infty$, or be bounded $[-a, a]$, depending on the domain of the variable. Consequently, this horizontal axis may correspond to the upper two quadrants of a Cartesian coordinate axis when the bar chart uses only the first quadrant (of positive values). Contiguous intervals are placed next to each other, reflecting the continuous nature of the variable.

On the vertical or ordinate axis, the heights of the intervals are positioned. The height is determined by dividing the relative frequency by the width of the interval: $h_i = f_i / a_i$. These representations constitute the density histogram. While resembling a bar chart, the bars are connected, the X-axis values have a mathematical meaning, and their areas correspond to the relative frequency of the interval.

The most common practice is for the intervals to have the same width, a determination guided by empirical rules such as the square root, Sturges [6], Doane [7], Scott [8], Freedman and Diaconis [9], and the Rice University rule [1]. Another option is to maintain homogeneity in density or frequency within each interval while allowing variable width. This approach is recommended when utilizing Pearson's chi-square test to assess the goodness-of-fit between the empirical distribution and a theoretical distribution [15]. Additionally, various algorithms exist for minimizing integrated mean square error [16], hazard function [17], or entropy [18].

There exists a variant of the histogram that sees more general usage when dealing with datasets featuring numerous values that have been grouped into class intervals. This variant is known as a frequency histogram, wherein the intervals of values share the same width and are positioned adjacent to each other. For each interval, a rectangle is drawn to a height corresponding to its absolute or simple relative frequency. In this case, the area under the curve is not equal to the frequency of the interval, as in the density histogram, but is a value proportional to it. Consequently, it resembles a bar chart where the bars are connected to each other. This approach ensures the preservation of the data's nature, avoiding the imposition of continuity on a discrete quantitative variable or the forced attribution of a quantitative and continuous character to an ordinal variable. The study uses this variant of the histogram for data analysis in order to facilitate its extrapolation to studies in the field of Psychology and related sciences.

3. Class Intervals When Tabulating and Graphing Sample Data

When dealing with a continuous quantitative variable and aiming to construct a frequency table, it is necessary to establish class intervals for conducting the frequency count. Once the table is defined, the data can be visually represented through a histogram. Class intervals are contiguous and non-overlapping ranges of values for the variable that adhere to two key principles: ordering and completeness. The ordering principle ensures the intervals are continuous and consecutive, while the completeness principle ensures that all sample data falls within a single interval, maintaining exclusivity.

During the creation of the table, the challenge arises in determining the number of class intervals (k) and deciding whether the intervals will have the same width (w) or not (w_i; $i = 1, 2, \ldots, k$). Here, the width of the class interval is defined as the difference between its upper limit (UL_i) and lower limit (LL_i): $w_i = UL_i - LL_i$.

It is recommended that the width of the intervals be constant (w), except for the two extreme intervals when long tails are generated by outliers. In this case, the width of these two extreme intervals can be much larger to encompass these low-frequency data points that are far away from the others.

There are several automatically applied or programmable rules for defining the number of class intervals, which can be classified into three groups [19] [20]. One group of rules starts by defining the constant amplitude, represented by a positive real number (w), and subsequently determines the number of intervals (k), which is a natural number. Within this group are rules designed to minimize an argument, typically either the integrated mean square error or the loss function [21]. Another set of rules begins by specifying the number of intervals (k), and from there determines the constant amplitude (w). In both groups of rules, the frequency or amount of data per interval (n_i) is variable. A third set of rules starts by setting the number of class intervals (k), and then establishes the homogeneous density or constant amount of data per interval (n), whereby the amplitude remains variable (w_i). This last strategy is employed to optimize the power of the goodness-of-fit chi-square test [15], and is not developed in this paper.

3.1. From the Constant Amplitude to the Number of Class Intervals

Within the first group, where rules define the constant amplitude (w) and subsequently determine the number of class intervals (k), the rules proposed by Scott [8] and Freedman and Diaconis [9] stand out for their simplicity and analytical formulation. Each of them is outlined below.

3.1.1. Scott's Rule (1979)

This rule assumes that the variable X follows a normal distribution and is grounded in the minimization of the integrated mean square error [8]. The am-

plitude (w) is derived from the quotient between 3.49 times the sample standard deviation (with the Bessel correction) and the cube root of the sample size (Equation (1)).

$$\begin{aligned} &\mathrm{x} = \left\{x_i\right\}_{i=1}^{n} = \left\{x_1, x_2, \cdots, x_n\right\} \subseteq \mathrm{X} \\ &w = \frac{3.49 \times s_{n-1}}{\sqrt[3]{n}} = \frac{3.49 \times \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}}{\sqrt[3]{n}} \\ &\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \end{aligned} \tag{1}$$

The number of intervals (k) is determined by rounding up the quotient of the total range (R) divided by the amplitude (w). See Equation (2).

$$k = \left\lceil \frac{R}{w} \right\rceil = \left\lceil \frac{\max\left(\left\{x_i\right\}_{i=1}^{n}\right) - \min\left(\left\{x_i\right\}_{i=1}^{n}\right)}{\frac{3.49 \times s_{n-1}}{\sqrt[3]{n}}} \right\rceil \tag{2}$$

Another option is to recalculate the constant amplitude (a_c) so that the k class intervals fall within the total range of the sample (R), inclusive of its minimum and maximum values but not exceeding them. Once the number of class intervals (k) is determined, the constant amplitude of the intervals is adjusted to match the sample range: $a_c = R/k$.

3.1.2. Freedman-Diaconis Rule (1981)

Freedman and Diaconis [9] make no assumptions about the distribution. This rule arises as a modification of Scott's rule designed to enhance robustness to outliers. It is derived from the quotient of twice the interquartile range and the cube root of the sample size (Equation (3)).

$$w = \frac{2R_{IQ}}{\sqrt[3]{n}} = \frac{2\left(q_{0.75} - q_{0.25}\right)}{\sqrt[3]{n}} \tag{3}$$

$R_{IQ} = q_{0.75} - q_{0.25}$ = interquartile range.

$q_{0.75}$ = third sample quartile or 0.75 order quantile.

$q_{0.25}$ = first sample quartile or 0.25 order quantile.

Optionally, the interval amplitude obtained (a) can be adjusted to remain within the sample range, encompassing both the minimum and maximum values without exceeding them. The adjusted amplitude (a_c) would be the quotient of the range and the number of class intervals: $a_c = R/k$.

To calculate the third and first quartiles, and subsequently obtain the interquartile range, one can utilize an interpolation rule based on the expected value or mean of the i-th order statistic from samples of size n drawn randomly from a continuous uniform distribution $U[0, 1]$. This order statistic follows a beta distribution with shape parameters: $\alpha = i$ and $\beta = n + 1 - i$ [22], whose mean is: $\mu = \alpha/(\alpha + \beta) = i/(n + 1)$. This rule is often referenced as rule 6 in R [23]. The SPSS program calculates sample quantiles using this rule [24]. When applying the functions CUARTIL.EXC and PERCENTIL.EXC in the Excel program, which

exclude quantile orders 0 and 1 from the calculation, this rule is employed [25].

When dealing with a normal sample distribution, the application of rule 9 is recommended in R [23]. In general, Hyndman and Fan [26] suggest utilizing rule 8 in the R program, which relies on the median of the distribution of the i-th order statistic in random samples of size n drawn from a standard continuous uniform distribution. This distribution follows the beta distribution mentioned earlier in the previous paragraph. The PERCENTILE_EXC function from the Real Statistics Resource Pack package facilitates the calculation of sample quantiles using these rules [25]. It's worth noting that this package also offers the Harrell-Davis robust (distribution-free) procedure for estimating quantiles. This procedure can be particularly useful with bimodal data, such as samples from the arcsine distribution, as well as very heavy-tailed symmetric distributions like the Cauchy, and asymmetric distributions such as the lognormal [27].

In both rules, after determining the constant or homogeneous amplitude, the process of constructing class intervals begins with identifying the minimum value in the sample. The amplitude is added to this minimum value, defining the lower limit of the first class interval. The upper limit of this interval becomes the lower limit of the next interval, and the amplitude is added again. This process continues until the maximum sample value in the k-th interval is included or exceeded. The number of class intervals (k) is calculated as the upwardly rounded quotient of the total sample range and the constant amplitude of the intervals: $k = \lceil R/w \rceil$. When using the Frequency function in the Excel program for frequency counting, the class interval is considered closed at its upper limit and open at its lower limit. However, an exception is made for the first interval, which includes the lower limit or sample minimum. This study utilizes Excel version 2021 as the software program.

3.2. From the Constant Amplitude to the Number of Class Intervals

Within the second group, which defines k class intervals and, from k, obtains the constant amplitude w, four rules stand out: square root [5], Sturges [6], Doane [7], and Rice University [1] [2]. Each of them is defined below.

3.2.1. Square Root Rule

Karl Pearson introduced the square root rule in his book "The Grammar of Science", published in 1892 [5]. Hence, it is the oldest rule for determining the number of class intervals [14]. This rule is implemented in various statistical packages and is commonly used in data analysis [28]. It is particularly recommended for sample sizes smaller than 100 [1] [29]. The rule is based on partitioning the n sample data into k groups of approximately k elements [6] (Equation (4)).

$$n = \overbrace{k + k + \cdots + k}^{k \text{ times}} = k \times k = k^2 \tag{4}$$

By isolating the variable k in the Equation (4), we find that the number of class intervals is the square root of the sample size rounded up (Equation (5)).

$$k = \lceil \sqrt{n} \rceil \quad (5)$$

3.2.2. Sturges' Rule (1926)

Esta regla asume simetría en la distribución y se basa en la aproximación de la distribución binomial $B(n = k - 1, p = 1/2)$ a la distribución normal cuando k tiende a infinito. Se sugiere para tamaños de muestra de 100 a 1000 ($100 \leq n \leq 1000$). Sin embrago, no está recomendada para tamaños muy grandes, pues resulta un número muy pequeño de intervalos de clase [30] [31].

This rule assumes symmetry in the distribution and relies on approximating the binomial distribution $B(n = k - 1, p = 1/2)$ to the normal distribution $N(\mu = (k - 1)/2, \sigma^2 = (k - 1)/4)$ when k tends to infinity. It is recommended for sample sizes ranging from 100 to 1000 ($100 \leq n \leq 1000$). However, it is not advisable for very large sample sizes, as it leads to a very small number of class intervals [30] [31].

Sturges [6] proposes that, if one has a sample of 16 data, taking as a model the binomial distribution $B(n = 4, p = 1/2)$, one could allocate these 16 elements into five groups or class intervals with the following frequencies: 1 data point for class 1, 4 data points for class 2, 6 for class 3, 4 for class 4, and 1 for class 5. When expressing 16 as a power of 2, the number of class intervals (k) corresponds to the exponent of this power increased by one unit: $16 = 24$, $k = 4 + 1 = 5$. If there are 32 elements, based on a binomial distribution $B(n = 5, p = 1/2)$, they would be allocated into six groups or class intervals with the following distribution: 1 data point for class 1, 5 data points for class 2, 10 for class 3, 10 for class 4, 5 for class 5, and 1 for class 6. When expressing 32 as a power of 2, the number of intervals (k) corresponds to the exponent plus one: $32 = 25$, $k = 5 + 1 = 6$.

In this approach, the n elements are distributed among k containers or bins, considering all distribution possibilities in the context of random sampling, without prior knowledge of the number of containers. For the first bin, the count includes the number of zero-element groups. In the second bin, it counts the number of one-element groups, and this pattern continues, increasing by one until the k-th bin is reached, where the elements remain in a single group. Consequently, the number of elements to form different groups is $k - 1$, with the order of the $k - 1$ elements being irrelevant, and no element repetition is allowed within each bin.

Expressed in arithmetic terms, the options for the first bin involve combinatorics without repetition of $k - 1$ elements taken in groups of 0 elements, resulting in one option. The options for the second bin involve combinatorics without repetition of $k - 1$ elements taken in groups of 1 element, resulting in k options. The options for the third bin involve combinatorics without repetition of $k - 1$ elements taken in groups of 2 elements, resulting in $[k \times (k - 1)]/2$ options. The options for the k-th bin involve combinatorics without repetition of $k - 1$ elements taken in groups of $k - 1$ elements, resulting in one option. The sum of these k combinatorics without repetition results in the total of n elements, which

is equivalent to 2 raised to the power of $k-1$ (Equation (6)).

$$\begin{aligned} n &= \sum_{i=0}^{k-1}\binom{k-1}{i} = \binom{k-1}{0}+\binom{k-1}{1}+\binom{k-1}{2}+\cdots+\binom{k-1}{k-1} \\ &= \binom{k-1}{0}1^0 1^{k-1}+\binom{k-1}{1}1^1 1^{k-2}+\cdots+\binom{k-1}{k-1}1^{k-1}1^0 \\ &= (1+1)^{k-1} = 2^{k-1} \end{aligned} \tag{6}$$

The value of k is isolated from the Equation (6) and rounded upwards (Equation (7)).

$$\begin{aligned} & n = 2^k \\ & \log_2(n) = \log_2\left(2^{k-1}\right) \\ & \log_2(n) = (k-1)\log_2(2) = k-1 \\ & k = 1+\lceil \log_2(n) \rceil = 1+\lceil \log(n)/\log(2) \rceil = 1+\lceil \ln(n)/\ln(2) \rceil \end{aligned} \tag{7}$$

The Sturges rule is widely used and recommended. However, as highlighted by Hyndman [32], it is not universally applicable. The rule performs inadequately with small and very large samples. It relies on an approximation of the symmetric binomial distribution to the normal distribution. Consequently, deviations from symmetry, such as a distribution with a heavy or elongated tail, or a marked departure from normality, for instance, a distribution with both tails heavily weighted or elongated, can negatively impact its accuracy.

3.2.3. Doane's Rule (1976)

It is a variant of the Sturges rule used in the presence of a skewed distribution [33]. The number of bins is determined by the formula provided in Equation (8):

$$\begin{aligned} k &= 1+\left\lceil \log_2(n)+\log_2\left(1+\frac{\left|\sqrt{b_1}\right|}{s_{\sqrt{b_1}}}\right)\right\rceil \\ &= 1+\left\lceil \log(n)/\log(2)+\log\left(1+\frac{\left|\sqrt{b_1}\right|}{s_{\sqrt{b_1}}}\right)\Big/\log(2)\right\rceil \\ &= 1+\left\lceil \log(n)/\log(2)+\ln\left(1+\frac{\left|\sqrt{b_1}\right|}{s_{\sqrt{b_1}}}\right)\Big/\ln(2)\right\rceil \end{aligned} \tag{8}$$

In this formula, the skewness measure is the skewness coefficient based on Karl Pearson's standardized third central moment [5]. It can be obtained from the sample mean of the cubed standardized values (with the standard deviation without the Bessel correction), as shown in Equation (9).

$$\begin{aligned} \sqrt{b_1} &= \frac{m_3}{m_2^{\frac{3}{2}}} = \frac{\sum_{i=1}^n (x_i-\bar{x})^3/n}{\left(\sqrt{\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n}}\right)^3} = \frac{\sum_{i=1}^n (x_i-\bar{x})^3}{n\left(\sqrt{s_n^2}\right)^3} \\ &= \frac{\sum_{i=1}^n (x_i-\bar{x})^3}{ns_n^3} = \frac{1}{n}\sum_{i=1}^n \left(\frac{x_i-\bar{x}}{s_n}\right)^3 = \frac{\sum_{i=1}^n z_i^3}{n} = \bar{z}^3 \end{aligned} \tag{9}$$

The standard deviation or error of Karl Pearson's [34] skewness coefficient ($\sqrt{b1}$) is calculated using the population formula proposed by Egon Sharpe Pearson [35] for a variable with a normal distribution. This formula relies solely on the sample size (n) and remains valid even when the distribution of the variable is not normal, provided it has a finite mean and variance, and as the sample size (n) tends to infinity (Equation (10)).

$$\sqrt{b_1} \sim N\left(\mu = 0, \sigma^2 = \frac{6(n-2)}{(n+1)(n+3)}\right) \tag{10}$$

3.2.4. Rice University Rule

It was developed in the statistics department at Rice University [1] [2] and, like the square root and Sturges rules, depends solely on the sample size. According to this rule, the number of class intervals (k) is determined by rounding upward twice the cube root of the sample size, as shown in Equation (11). This factor precisely corresponds to the denominator of the rules proposed by Scott [8] and Freedman and Diaconis [9].

$$k = 2 \times \left\lceil \sqrt[3]{n} \right\rceil = 2 \times \left\lceil n^{1/3} \right\rceil \tag{11}$$

It can be seen as a simplification of these rules when the homogeneous amplitude (w) is redefined to include the minimum and maximum values of the sample without exceeding them ($\lceil R/w \rceil = k_{aj}$; $w_{aj} = R/k_{aj}$) and a value approximating is given the range. This value is about seven times the sample standard deviation from the Scott's rule (Equation (12)) or eight times the semi-interquartile range from Freedman-Diaconis rule (Equation (13)). These broad ranges allow for the inclusion of extremely atypical cases, defined as those more than three standard deviations from the mean or more than three times the interquartile range from the median.

$$\mathrm{x} = \{x_i\}_{i=1}^{n} = \{x_1, x_2, \cdots, x_n\}$$

$$\begin{aligned} w_{aj} &= \left\lceil \frac{R(\mathrm{x})}{w_{FD}} \right\rceil = \left\lceil \frac{\max(\mathrm{x}) - \min(\mathrm{x})}{\frac{2\left[q_{0.75}(\mathrm{x}) - q_{0.25}(\mathrm{x})\right]}{\sqrt[3]{n}}} \right\rceil \approx \left\lceil \frac{4\left[q_{0.75}(\mathrm{x}) - q_{0.25}(\mathrm{x})\right]}{\frac{2\left[q_{0.75}(\mathrm{x}) - q_{0.25}(\mathrm{x})\right]}{\sqrt[3]{n}}} \right\rceil \\ &= \left\lceil \frac{8\left\lfloor \frac{\left[q_{0.75}(\mathrm{x}) - q_{0.25}(\mathrm{x})\right]}{2} \right\rfloor}{\frac{2\left[q_{0.75}(\mathrm{x}) - q_{0.25}(\mathrm{x})\right]}{\sqrt[3]{n}}} \right\rceil = 2\left\lceil \sqrt[3]{n} \right\rceil \end{aligned} \tag{12}$$

$$w_{aj} = \left\lceil \frac{R(\mathrm{x})}{w_{SCott}} \right\rceil \approx \left\lceil \frac{\max(\mathrm{x}) - \min(\mathrm{x})}{\frac{3.49 \times s_{n-1}(\mathrm{x})}{\sqrt[3]{n}}} \right\rceil = \left\lceil \frac{6.98 \times s_{n-1}(\mathrm{x})}{\frac{3.49 \times s_{n-1}(\mathrm{x})}{\sqrt[3]{n}}} \right\rceil = 2\left\lceil \sqrt[3]{n} \right\rceil \tag{13}$$

After determining the number of intervals (k) using one of these four rules, the constant amplitude (w) is obtained by dividing the sample range (numera-

tor), or the difference between the maximum value (minuend) and the minimum value (subtrahend) of the sample, by the number of intervals (denominator), as shown in Equation (14).

$$w = \frac{R}{k} = \frac{\max\left(\{x_i\}_{i=1}^{n}\right) - \min\left(\{x_i\}_{i=1}^{n}\right)}{k} \tag{14}$$

4. Method

On the one hand, 35 samples were randomly drawn from a standard continuous uniform distribution with seven different sizes (20, 35, 50, 50, 100, 100, 200, 500, and 1000), five samples for each size. The extraction was performed using the Excel random number generator.

$$\mathrm{u} = \{u_i\}_{i=1}^{n} \subset U[0,1]$$

Using the inverse transform sampling method, a total of 350 samples were generated from the initial set of 35 uniformly distributed samples. Seventy of these samples followed symmetric mesokurtic distributions. Specifically, 35 samples were drawn from a normal distribution with a location parameter $\mu = 5$ and a squared scale $\sigma^2 = 6.25$ (Equation (15)), and the other 35 samples were generated from a beta distribution with shape parameters $\alpha = 30$ and $\beta = 30$ (Equation (16)).

$$\Phi^{-1}(x_i) = 5 + 2.5\Phi^{-1}(u_i);\ X \sim N\left(\mu = 5, \sigma^2 = 6.25\right) \tag{15}$$

$$Q_X(u_i) = I_{u_i}^{-1}(\alpha = 30, \beta = 30);\ X \sim \mathrm{Beta}(\alpha = 30, \beta = 30) \tag{16}$$

Out of the total of 350 samples, 70 followed leptokurtic symmetric distributions, Specifically, 35 were drawn from the Laplace distribution with a location parameter $\mu = 5$ and a scale parameter $\beta = 2.5$ (Equation (17)), and the remaining 35 were obtained from the logistic distribution with a location parameter $x_0 = 5$ and a scale parameter $\gamma = 2.5$ (Equation (18)).

$$\begin{aligned} &Q_X(u_i) = 5 - 2.5 \times \mathrm{signo}(u_i - 0.5) \times \ln\left(1 - 2\left|u_i - 0.5\right|\right); \\ &X \sim \mathrm{Laplace}(\mu = 5, \beta = 2.5) \end{aligned} \tag{17}$$

$$Q_X(u_i) = 5 + 2.5 \times \ln\left(\frac{u_i}{1 - u_i}\right);\ X \sim \mathrm{Logistica}(\mu = 5, s = 2.5) \tag{18}$$

Out of the total of 350 samples, 70 followed symmetric platykurtic distributions. Specifically, 35 were drawn from the arcsine distribution with threshold parameters 0 and 1 (Equation (19)), and the remaining 35 from the semicircular distribution with a radius of 2 (Equation (20)). The standard arcsine distribution is equivalent to a beta distribution of shape parameters: $\alpha = 0.5$ and $\beta = 0.5$, whose quantile function corresponds to an inverse regularized incomplete beta function: I_{ui}^{-1} ($\alpha = 0.5$, $\beta = 0.5$), as seen in Equation (20) The semicircular distribution with parameter R (radius) corresponds to a beta distribution of shape parameters: $\alpha = 1.5$ and $\beta = 1.5$, once transformed: $X = 2 \times R \times (B - 1)$, where $B \sim$ Beta ($\alpha = 0.5$, $\beta = 0.5$) and $X \sim$ Arcsine ($R = 2$), as shown in Equation (21).

$$\begin{aligned} &Q_X\left(u_i\right)=I_{u_i}^{-1}\left(\alpha=0.5,\beta=0.5\right); \\ &X\sim\text{Arcoseno}\left(a=0,b=1\right)\equiv\text{Beta}\left(\alpha=0.5,\beta=0.5\right) \end{aligned} \tag{19}$$

$$Q_X\left(u_i\right)=4I_{u_i}^{-1}\left(\alpha=1.5,\beta=1.5\right)-2;\ X\sim\text{Semicircular}\left(R=2\right) \tag{20}$$

Out of the total of 350 samples, 70 followed distributions with skewness and platykurtosis. Specifically, 35 were drawn from the triangular distribution with parameters $a = b = 0$ and $c = 1$ (Equation (21)), and the remaining 35 from the PERT distribution with parameters $a = 1$, $b = 4$ and $c = 5$, where a is the minimum, b is the peak or mode and c is the maximum (Equation (22)).

$$X=1-\sqrt{1-u_i};\ X\sim\text{Triangular}\left(a=0,b=0,c=1\right) \tag{21}$$

$$\begin{aligned} &\alpha=1+4\frac{b-a}{c-a}=1+4\frac{4-1}{5-1}=4 \\ &\beta=1+4\frac{c-b}{c-a}=1+4\frac{5-4}{5-1}=2 \\ &Q_X\left(u_i\right)=1+4\times I_{u_i}^{-1}\left(\alpha=4,\beta=2\right);\ X\sim\text{PERT}\left(a=1,b=4,c=5\right) \end{aligned} \tag{22}$$

Finally, out of the total 350 samples, 70 followed distributions with skewness and leptokurtosis. Specifically, 35 were drawn from an exponential distribution with a rate parameter $\lambda = 1/2$ (Equation (23)), and the other 35 are obtained from a lognormal distribution with a location parameter $\mu_{\ln(X)} = 0$ and a scale parameter $\sigma_{\ln(X)} = 0.25$ (Equation (24)).

$$Q_X\left(u_i\right)=-2\ln\left(1-u_i\right);\ X\sim\exp\left(\lambda=0.5\right) \tag{23}$$

$$Q_X\left(u_i\right)=\mathrm{e}^{\frac{\sqrt{2}}{4}\text{erf}^{-1}\left(2u_i-1\right)}=\mathrm{e}^{0.25\Phi^{-1}\left(u_i\right)};\ X\sim\log N\left(\mu_{\ln(X)}=0,\sigma_{\ln(X)}=0.25\right) \tag{24}$$

The six empirical rules for determining the number (k) and width (w) of bins were applied to the 350 random samples, resulting in 2100 empirical histograms. The probability of the k class intervals (I) of each empirical histogram was calculated using the cumulative distribution function of the corresponding distribution to generate the theoretical histogram. The difference between the empirical and theoretical histograms was measured using the average Euclidean distance between the relative frequency $f_n(I_i)$ and the expected probability $p_X(I_i)$ for each bin ($i=1,2,\cdots,k$). This difference is referred to as the Average Discrepancy (AD) and is shown in Equation (25).

$$\begin{aligned} &\text{AD}=\sqrt{\frac{\sum_{i=1}^{k}\left[f_n\left(I_i\right)-p_X\left(I_i\right)\right]^2}{k}} \\ &p_X\left(I_i\right)=\int_{LL}^{UL}f_X\left(x\right)\text{d}x;\ I_i=\left(LL_i,UL_i\right];\ i=1,2,\cdots,k \end{aligned} \tag{25}$$

On the other hand, a sample of 1000 uniformly distributed data with perfect symmetry was generated in the interval (0, 1), starting at 0.001 and ending at 0.999, with a data spacing of 0.000998. Ten theoretical samples were created using the corresponding density functions to represent their density curves. Each density curve, confined within the sample range, was used to visually assess (by a single judge) whether the empirical histogram reproduces the curve corres-

ponding to the distribution, using four ordered categories: 1 = not at all, 2 = a little, 3 = quite a lot, and 4 = completely, referred to as the Recognition Level (RL).

An index, termed the Accuracy Index (AI), was established using the Average Discrepancy (AD) and Recognition Level (RL). Initially, the difference between empirical and theoretical histograms (AD) is transformed into scores ranging from 1 to 4, where 4 signifies maximum accuracy and 1 denotes minimum inaccuracy among the 2100 samples (5 samples per triple condition: 10 distribution types × 7 sample sizes × 6 rules), as shown in Equation (26).

$$\begin{aligned} D_T &= 1+3\left(1-\frac{\mathrm{AD}-\min(\mathrm{AD})}{\max(\mathrm{AD})-\min(\mathrm{AD})}\right) \\ &= 1+3\left(1-\frac{\mathrm{AD}-\min\left(\{ad_i\}_{I=1}^{2100}\right)}{\max\left(\{d_i\}_{I=1}^{2100}\right)-\min\left(\{d_i\}_{I=1}^{2100}\right)}\right) \\ &= 1+3\left(1-\frac{\mathrm{AD}-0.00251545}{0.21781641-0.00251545}\right) \end{aligned} \tag{26}$$

Subsequently, AD and RL are summed, resulting in a random variable (D_T) with a potential range of 2 to 8 (Equation (27)). This combined variable is then transformed into the Accuracy Index (AI) with values ranging from 0 to 100, where 0 signifies total inaccuracy and 100 indicates total accuracy (Equation (28)).

$$S = D_T + \mathrm{RL} \tag{27}$$

$$\mathrm{AI} = 100\times\frac{S-\min(S)}{\max(S)-\min(S)} = 100\times\frac{S-2}{8-2} = 100\times\frac{D_T+GR-2}{6} \tag{28}$$

The normality of the distributions for AD, RL, and AI was assessed using Shapiro-Francia W' test [36] and K^2 test [37]. Means for each of the three variables were compared across three factors: rule (six levels), sample size (seven levels), and distribution (ten levels). This was done through a three-factor aligned rank transformation analysis of variance [38]. Pairwise comparisons were conducted using Fisher's least significant difference test [39] with Holm-Bonferroni correction applied to control the family rate error [40]. The effect size was estimated using the partial eta coefficient: $\eta^2 = F\times df_1/(F\times fd_1 + df_2)$. Interpretation of η^2 values was as follows: less than 0.02 indicates a very small effect size, between 0.02 and 0.129 is considered small, between 0.13 and 0.259 is medium, and greater than or equal to 0.26 is considered large [41] [42]. Data analyses were performed using EXCEL 2023, IBM SPSS Statistics 29, and R 4.3. The significance level was set at 0.05.

5. Results

5.1. Average Discrepancy (AD)

The distribution of the average discrepancy between the observed and expected histograms (AD) showed positive skewness or right-tailedness (g_1 = 1.237, 95%

CI [1.133, 1.342]) and platykurtosis or shortened tails with respect to the shoulders (g_2 = −0.227, 95% *CI* [−0.332, −0.122]). This indicates a departure from a normal distribution (Shapiro-Francia W' statistic = 0.885, p < 0.001; D'Agostino-Berlanger-D'Agostino K^2 statistic = 433.868, p < 0.001). See Figure 1.

In comparing the means of AD across the factors of Distribution type (D), Sample size (n), and Rule to determine the number and width of class intervals (Rule) using aligned rank transformation analysis of variance, the main effect of all three factors was found to be significant (Table 1).

The effect size of Sample size on AD was large (Table 1). The larger the sample size, the smaller the AD (Figure 2). The linear correlation, as indicated by Spearman's coefficient, between Sample size (n) and ranks (aligned with respect to the Sample size) for AD is very high: r_S = −0.925, p < 0.001.

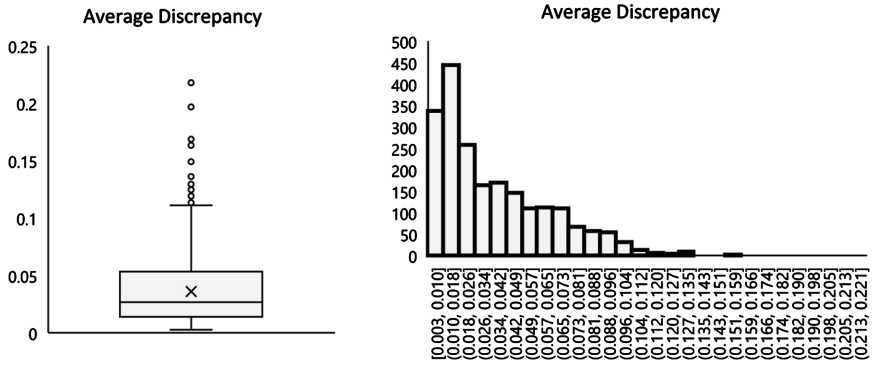

Figure 1. Box plot (left graph) and histogram with the density curve by Kernel (Gaussian) estimation superimposed (right graph) of AD. The width and number of class intervals was established by the Friedman-Diaconis rule.

Table 1. Aligned rank transform ANOVA for the average discrepancy between the observed and expected frequencies of class intervals.

Factors	*SS*	df_1	df_2	*MS*	*F*	*p-value*	$_p\eta^2$	*ES*
D	62706824.74	9	1680	6967424.97	17.37	<0.001	0.085	s
n	678789373.9	6	1680	113131562.32	2097.68	<0.001	0.882	l
Rule	34437330.12	5	1680	6887466.02	16.46	<0.001	0.047	s
D × n	41067732.2	54	1680	760513.56	1.82	<0.001	0.055	s
D × Rule	1502654.22	45	1680	33392.32	0.08	1	0.002	vs
n × Rule	44323209.93	30	1680	1477440.33	3.57	<0.001	0.060	s
D × n × Rule	98998460.43	270	1680	366660.96	0.95	0.686	0.133	m

Note. Factors: D = Distribution type, n = Sample size, Rule = Rule to determine the number and width of class intervals, × = interaction between factors. *SS* = sum of squares, *df* = degrees of freedom, *MS* = mean squares, *F* = testing statistic, *p*-value = right-tailed probability in a Snedecor-Fisher F distribution with degrees of freedom df_1 and df_2, $_p\eta^2$ = partial eta squared as effect size estimator, *ES* = effect size (Cohen, 1992): l = large, m = medium, s = small, and vs = very small.

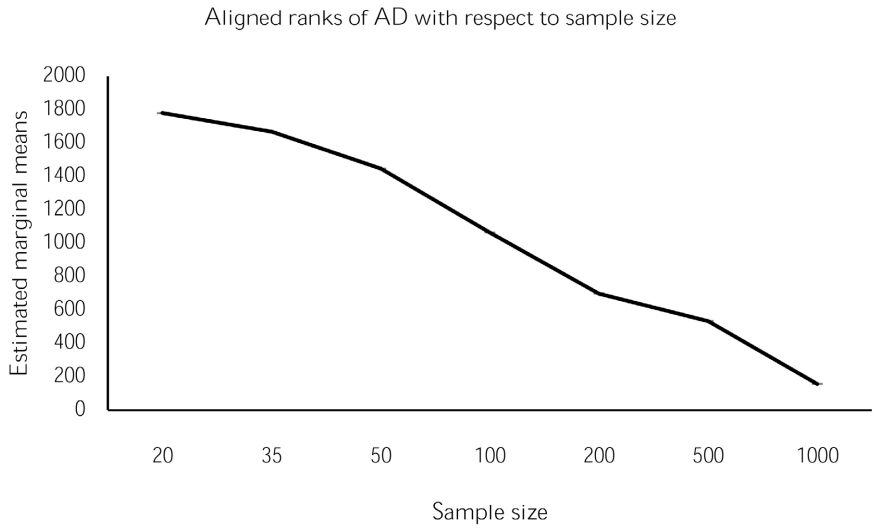

Figure 2. Plot of means conditioned on Sample Size of the ranks (aligned with respect to the Sample Size) for AD.

The size of the effect of the distribution on AD was small (**Table 1**). The lowest mean rank (aligned with respect to the distribution) for the AD appeared with the exponential distribution, while the highest mean rank was observed with the arcsine distribution.

Out of 45 pairwise comparisons, 25 (55.6%) were found to be significant after Holm-Bonferroni correction. The normal distribution exhibited a significantly lower mean rank than the arcsine and a higher mean rank than the exponential distribution. The beta distribution showed a lower mean rank than the arcsine and a higher mean rank than the exponential, Laplace, logistic, and lognormal distributions. The Laplace distribution had a significantly lower mean rank than the arcsine, semicircular, triangular, and PERT distributions. The logistic distribution also had a lower mean rank than the arcsine, semicircular, triangular, and PERT distributions, but a higher mean rank than the exponential. The arcsine, semicircular, triangular, and PERT distributions had higher mean ranks than the exponential and lognormal. The exponential distribution had a lower mean rank than the lognormal (**Figure 3**).

The effect size of the Rule on AD was small (**Table 1**). Out of the 15 differences, 10 (66.7%) were found to be significant after Holm-Bonferroni correction. The square root rule and Doane had the lowest mean ranks (aligned with respect to the Rule) for AD, while Scott's had the highest mean rank. The average ranks of the square root and Rice rules were significantly lower than those of the Scott, Friedman-Diaconis, and Sturges rules. The mean rank of the Sturges rule was significantly lower than that of the Scott rule but higher than that of the Doane rule. The mean rank of the Doane rule was significantly lower than those of the Scott and Friedman rules. The mean ranks of the square root, Rice, and Doane rules were equivalent. The Friedman-Diaconis rule had an average rank equivalent to that of Sturges and Scott (**Figure 4**).

Figure 3. Plot of means conditional on the Distribution type of the ranks (aligned with respect to the Distribution type) for AD.

Figure 4. Plot of means conditional on the Rule of the ranks (aligned with respect to the Rule) for AD.

The interaction of Sample size with Distribution type and Rule had significant effects on AD with small effect sizes, but not the interaction between Distribution type and Rule (**Figure 5**). Neither was the third-order interaction significant.

Concerning the interaction between Distribution type and Sample size, increasing sample size favors the arcsine distribution more due to the downward trend of its mean in aligned ranks and disfavors the Laplace distribution because of its upward trend (**Figure 6**). Regarding the interaction between Rule and Sample size, Scott's and Friedman's rules are the most favored by increasing sample size. On the contrary, Doane's rule is the most disadvantaged. Sturges' rule is also not helped by increasing sample size. The square root and Sturges rules do not exhibit a clear trend (**Figure 7**).

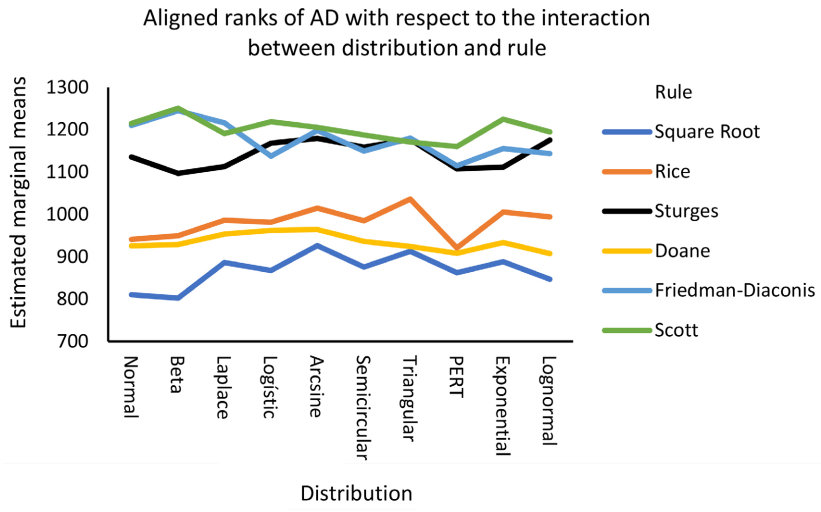

Figure 5. Plot of means conditional on the Distribution type and Rule of the ranks (aligned with respect to the interaction between the Distribution type and Rule) for AD.

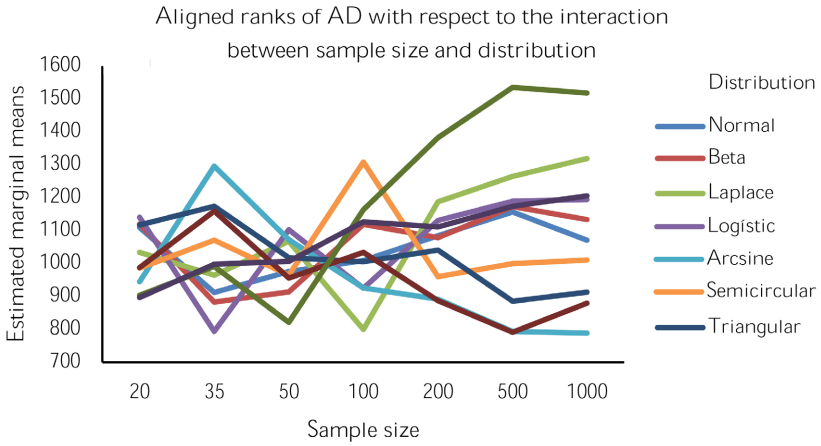

Figure 6. Plot of means conditional on the Distribution type and Sample size of the ranks (aligned with respect to the interaction between the Distribution type and Sample size) for AD.

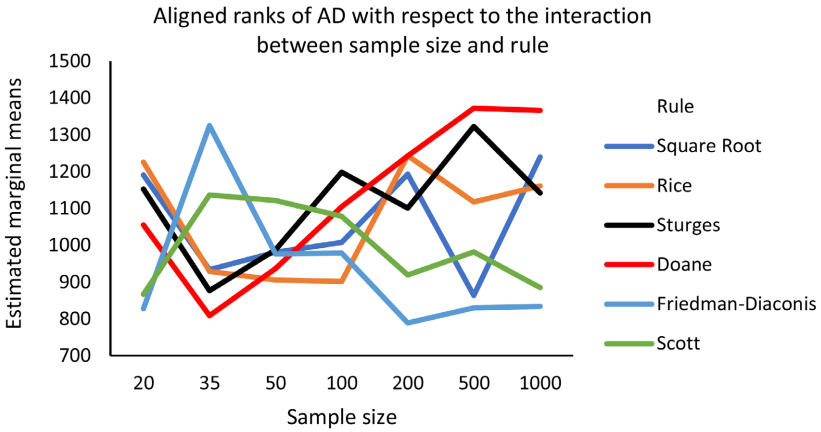

Figure 7. Plot of means conditional on the Rule and Sample size of the ranks (aligned with respect to the interaction between the Rule and Sample size) for AD.

5.2. Recognition Level (RL)

The distribution of the ordinal variable of Recognition Level (RL) showed negative asymmetry (Bowley's coefficient of skewness = −1, bias = 0, std error = 0, 99% percentile bootstrap confidence interval [−1, −1]; percentile coefficient of skewness = −0.333, bias = 0.0013, std error = 0.019, 99% percentile bootstrap confidence interval [−0.3333, −0.2007]) and platykurtosis or shortened tails in relation to the shoulders (Percentile kurtosis = −0.096, bias = 0.0008, std error = 0.0079, 99% percentile bootstrap confidence interval [−0.0965, −0.0132). The confidence interval is widened to 99% because the standard error is so small that it causes the upper and lower limits of the 95% interval to coincide. See **Figure 8**.

By comparing RL among the factors of Distribution type (D), Sample size (n), and the Rule to determine the number and width of class intervals (Rule) using an aligned rank transformation analysis of variance, the main effect of the three factors was significant (**Table 2**).

The effect size of the Sample size on RL was large (**Table 2**). The larger the sample size, the smaller the RL. The linear correlation, as indicated by the Spearman coefficient between the Sample size and the ranks (aligned with respect to the distribution) for RL was high, $r_S = 0.638$, $p < 0.001$ (**Figure 9**). The correlation of Sample size was significantly smaller with RL than with AD: Rosner-Glynn transformation [43]: $r_S(n, \text{AD}) = -0.920$, Rosner-Glynn transformation [43]: $r(\text{n, AD}) = -0.877$; $r_S(n, \text{RL}) = 0.482$, Rosner-Glynn transformation [43]: $r(n, \text{RL}) = 0.479$; $r_S(\text{AD, RL}) = -0.532$; Rosner-Glynn transformation [43]: −0.521; Meng-Rosenthal-Rubin z statistic [44] = −49.449, p-value ≤ 0.001; $r(\text{n, AD}) - r(\text{n, RL}) = -1.356$; 95% CI (−1.958, −1.809); effect size: $d = \sqrt{(n-3)} \times |z| = 1710.835$.

Table 2. Aligned rank transform ANOVA for the recognition level between the observed and expected frequencies of class intervals.

Factors	SS	df_1	df_2	MS	F	p-value	${}_p\eta^2$	ES
D	281535821.3	9	1680	31281757.92	108.62	<0.001	0.368	l
n	334602801.2	6	1680	55767133.53	219.57	<0.001	0.440	l
Rule	10568611.27	5	1680	2113722.25	4.77	<0.001	0.014	vs
D × n	110663206.5	54	1680	2049318.64	5.27	<0.001	0.145	m
D × Rule	103832566.9	45	1680	2307390.38	5.91	<0.001	0.137	m
n × Rule	48556148.91	30	1680	1618538.30	3.82	<0.001	0.064	s
D × n × Rule	100622795.5	270	1680	372677.02	0.94	0.739	0.131	m

Note. Factors: D = Distribution type, n = Sample size, Rule = Rule to determine the number and width of class intervals, × = interaction between factors. SS = sum of squares, df = degrees of freedom, MS = mean squares, F = testing statistic, p-value = right-tailed probability in a Snedecor-Fisher F distribution with degrees of freedom df_1 and df_2, ${}_p\eta^2$ = partial eta squared as effect size estimator, ES = effect size (Cohen, 1992): l = large, m = medium, s = small, and vs = very small.

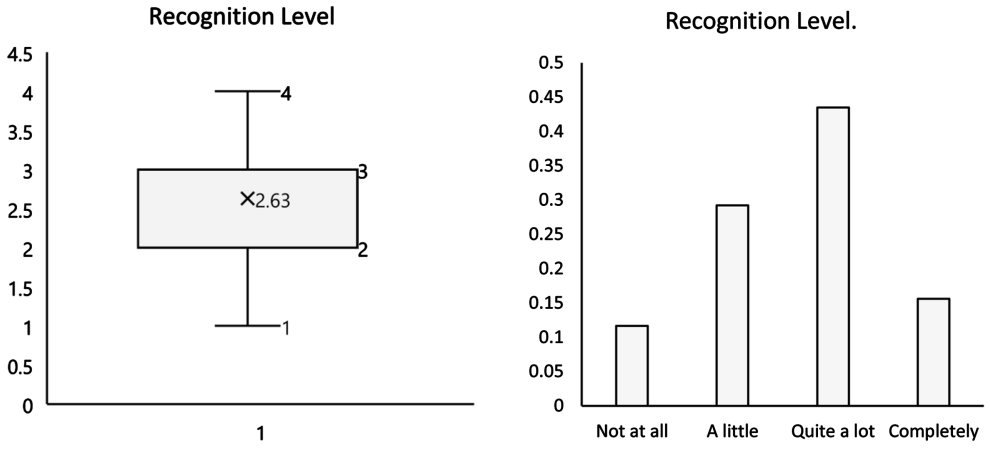

Figure 8. Box plot (left graph) and bar chart (right graph) for recognition level.

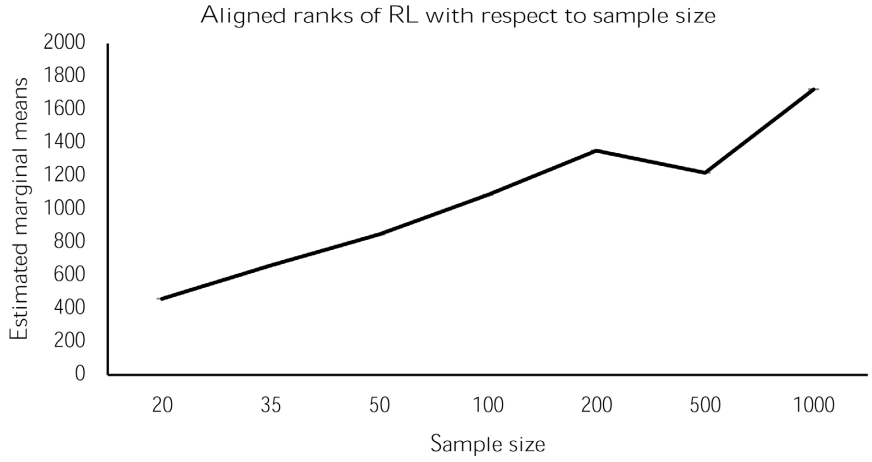

Figure 9. Plot of means conditional on the Sample size of the ranks (aligned with respect to the Sample size) for RL.

The effect size of distribution on RL was large (Table 2). The smallest mean ranks (aligned with respect to the Distribution type) for RL appears with the semicircular distribution, and the highest with the exponential distribution. Out of 45 comparisons, 33 (73.3%) were significant, and 12 (26.7%) were not, after Holm-Bonferroni correction. The mean ranks for RL were equivalent between the normal, beta, and logistic distributions. The mean ranks between the Laplace, arcsine, triangular, and PERT distributions were also equivalent. In turn, the mean rank of the lognormal distribution was equivalent to those of the Laplace, arcsine, and triangular distributions (Figure 10).

The effect size of the Rule on RL was very small (Table 2). The square root and Friedman-Diaconis rules had the lowest mean ranks (aligned with respect to the Rule) for RL, and Scott's rule had the highest mean rank. Out of the 15 differences, 3 (20%) were significant after the Holm-Bonferroni correction. The mean rank of the Scott rule was significantly higher than those of the square root and Friedman-Diaconis rules, and the mean rank of the Sturges rule was significantly higher than that of the square root rule (Figure 11).

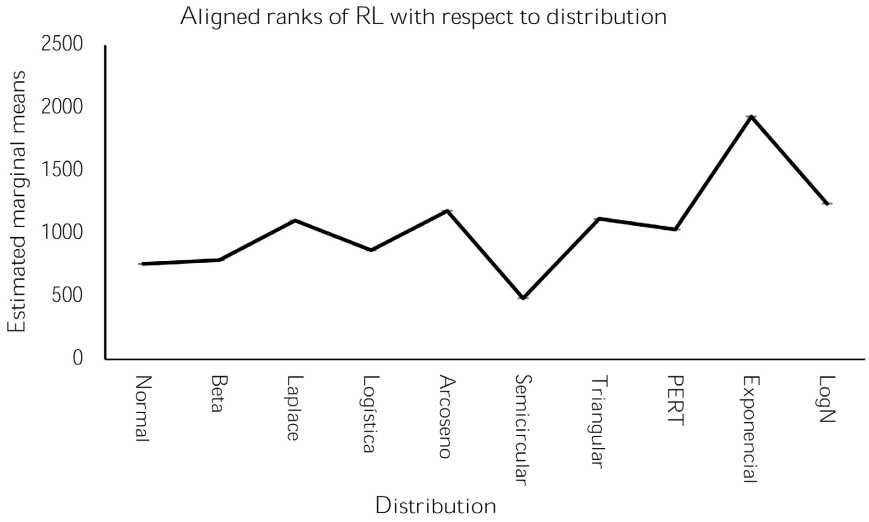

Figure 10. Plot of means conditional on the Distribution type of the ranks (aligned with respect to the Distribution type) for RL.

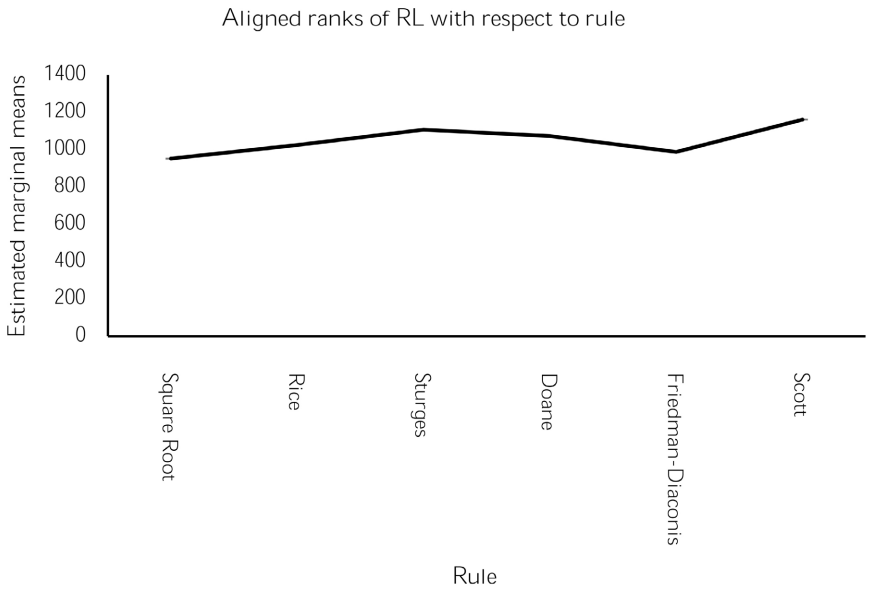

Figure 11. Plot of means conditional on the Rule of the ranks (aligned with respect to the Rule) for RL.

In RL, the second-order interaction effects were significant, but the third-order one was not. The effect size of the interactions of the Distribution type with the Sample size and the Rule was medium, and the size of the interaction between the Sample size and the Rule was small (**Table 2**).

Regarding the interaction between the Distribution type and the Rule, the arcsine distribution achieves the best mean ranks in RL with the square root, Rice, Sturges, and Doane rules, and the lowest ranks with the Friedman-Diaconis and Scott rules. The rules of Scott and Friedman-Diaconis favor the triangular distribution. This last rule also stands out with the semicircular distribution and PERT (**Figure 12**).

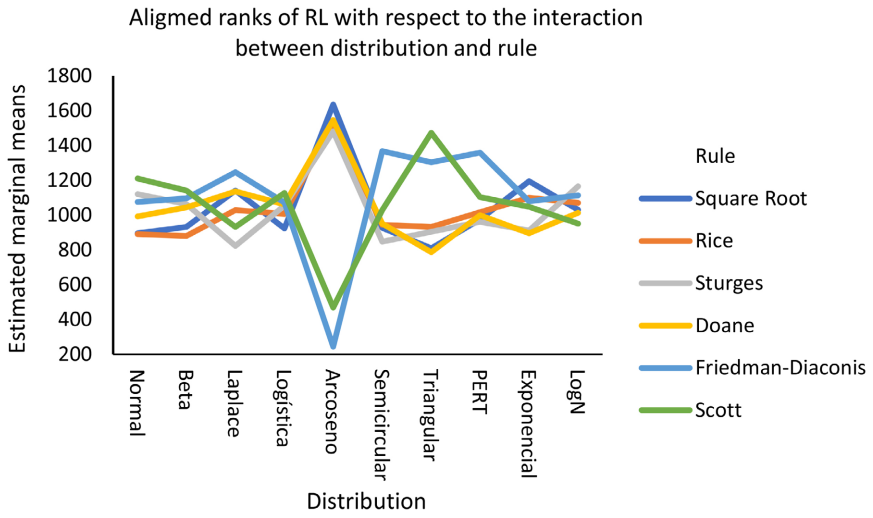

Figure 12. Plot of means conditional on the Rule and Distribution type of the ranks (aligned with respect to the interaction between the Rule and Distribution type) for RL.

Regarding the interaction between Distribution type and Sample size, the smallest sample sizes (20 and 35) favor the exponential distribution, while the size of 1000 has a detrimental effect on it. Sample sizes of 20, 50, and 1000 generate notable differences. With a size of 20, the Laplace distribution has the lowest mean rank, and the triangular and exponential distributions have the highest. With a size of 50, the Laplace distribution achieves the highest mean rank, and the exponential and semicircular distributions have the lowest. With 1000, the exponential and arcsine distributions have the lowest mean ranks, while the semicircular has the highest (**Figure 13**).

Regarding the interaction between the Rule and the Sample size, an increase in the sample size favors the Doane, Rice, and Sturges rules due to the upward trend of the curve. Conversely, the rules of Scott, Friedman-Diaconis, and the square root are detrimental due to the downward trend of the curve (**Figure 14**).

The correlation between AD and RL was significant, negative, and with a high strength of association (r_S = −0.532, *IC* al 95% [−0.568, −0.495], t [2098] = −28.753, $p < 0.001$).

5.3. Accuracy Index (AI)

Because the variable exhibits negative asymmetry (g_1 = −0.451, *IC* al 95% [−0.556, −0.346]), where the left tail is longer than the right, and platykurtosis (g_2 = −0.297, *IC* al 95% [−0.506, −0.088]), indicating shortened tails with respect to the shoulders, it deviates from normality (Shapiro-Francia W' statitic = 0.960, $p < 0.001$). See **Figure 15**.

By comparing the means of the Accuracy Index across the factors of Distribution type (D), Sample size (n), and the Rule to determine the number and width of class intervals (Rule) using an aligned rank transformation analysis of variance, the main effect of the three factors was significant (**Table 3**).

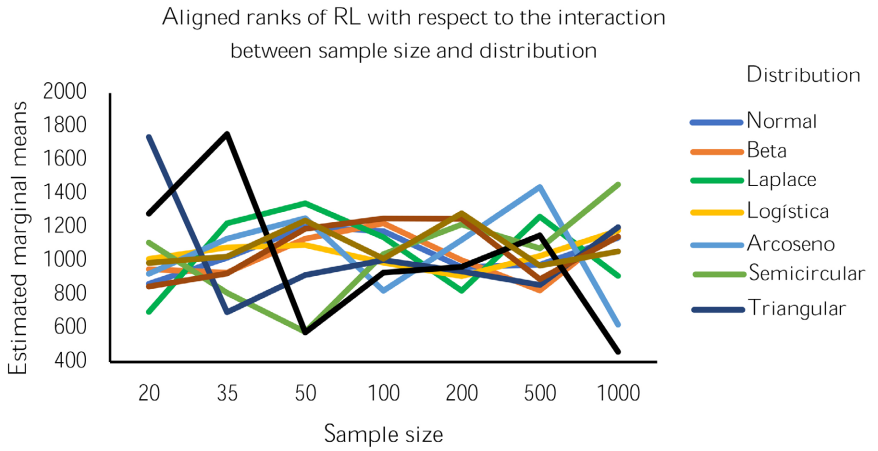

Figure 13. Plot of means conditional on the Sample size and Distribution type of the ranks (aligned with respect to the interaction between the Sample size and Distribution type) for RL.

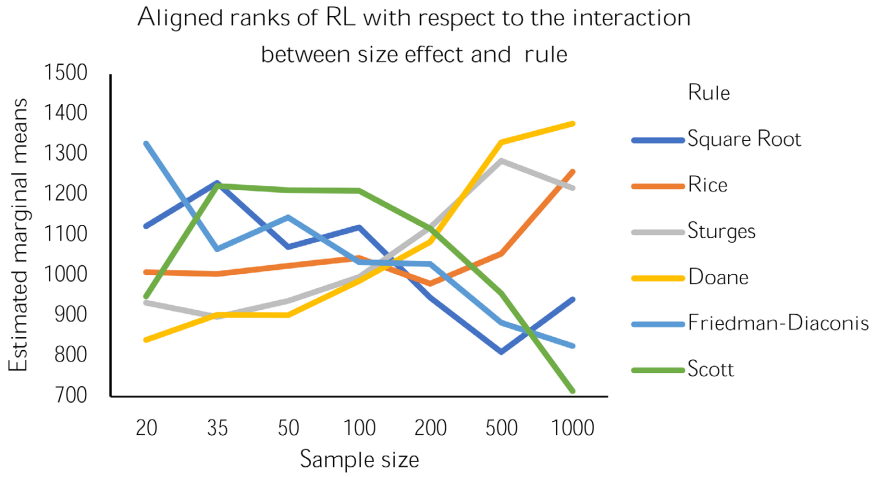

Figure 14. Plot of means conditional on the Rule and Sample size of the ranks (aligned with respect to the interaction between the Rule and Sample size) for RL.

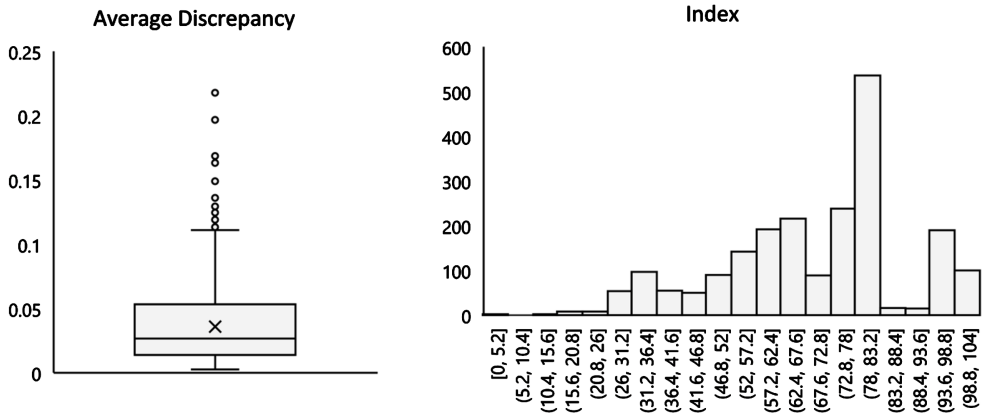

Figure 15. Box plot (left graph) and histogram with the density curve by Kernel (Gaussian) estimation superimposed (right graph) of Accuracy Index. The width and number of class intervals was established by the Friedman-Diaconis rule.

Table 3. Aligned rank transform ANOVA for the accuracy index between the observed and expected frequencies of class intervals.

Factors	*SS*	df_1	df_2	*MS*	*F*	*p-value*	${}_p\eta^2$	*ES*
D	257913572.4	9	1680	28657063.60	94.74	0.000	0.337	l
n	517065512.9	6	1680	86177585.49	581.71	0.000	0.675	l
Rule	7153980.931	5	1680	1430796.19	3.21	0.007	0.009	vs
D × *n*	94580933.65	54	1680	1751498.77	4.40	0.000	0.124	s
D × Rule	99671808.35	45	1680	2214929.07	5.66	0.000	0.132	m
n × Rule	37865991.61	30	1680	1262199.72	2.93	0.000	0.050	s
D × *n* × Rule	97261252.08	270	1680	360226.86	0.91	0.851	0.127	s

Note. Factors: D = Distribution type, *n* = Sample size, Rule = Rule to determine the number and width of class intervals, × = interaction between factors. *SS* = sum of squares, *df* = degrees of freedom, *MS* = mean squares, *F* = testing statistic, *p*-value = right-tailed probability in a Snedecor-Fisher F distribution with degrees of freedom df_1 and df_2, ${}_p\eta^2$ = partial eta squared as effect size estimator, *ES* = effect size (Cohen, 1992): l = large, m = medium, s = small, and vs = very small.

The effect size of the Sample size on Accuracy Index was large. The larger the sample size, the higher the accuracy rate (Table 3). The linear correlation, as measured by the Spearman coefficient between the Sample size and the ranks (aligned in relation to Sample size) for the Accuracy Index, is positive with a very strong strength of association, $r_S = 0.807$, $p < 0.001$ (Figure 16).

The effect size of the distribution on the Accuracy Index was large (Table 3). The highest mean rank (aligned with respect to Distribution type) for the Accuracy Index appears with the exponential distribution, and the lowest with the semicircular distribution. In both cases, there is a significant difference compared to the other distributions. Out of 45 comparisons, 32 (71.1%) were significant after Holm-Bonferroni correction. The 13 equivalences (28.9%) were observed between the normal distribution and the beta and logistic distributions, between the beta and the logistic distributions, between the Laplace distribution and the arcsine, triangular, PERT, and lognormal distributions, between the logistic and the PERT distributions, between the arcsine distribution and the triangular, PERT, and lognormal distributions, as well as between the triangular distribution and PERT and lognormal distributions (Figure 17).

The effect size of the Rule on the Accuracy Index was very small (Table 3). Out of the 15 differences, 6 (40%) were significant, and 9 (60%) were not. The mean ranks (aligned with respect to the Rule) for the Accuracy Index corresponding to the square root and Friedman-Diaconis rules were lower than the mean ranks corresponding to the Sturges, Doane, and Scott rules. However, none were significant after the Holm-Bonferroni correction (Figure 18).

The second-order interaction effects on the Accuracy Index were significant. The interaction between Distribution type and Rule had a medium effect size, and the other two interactions had a small effect size, but the third-order interaction was not significant (Table 3).

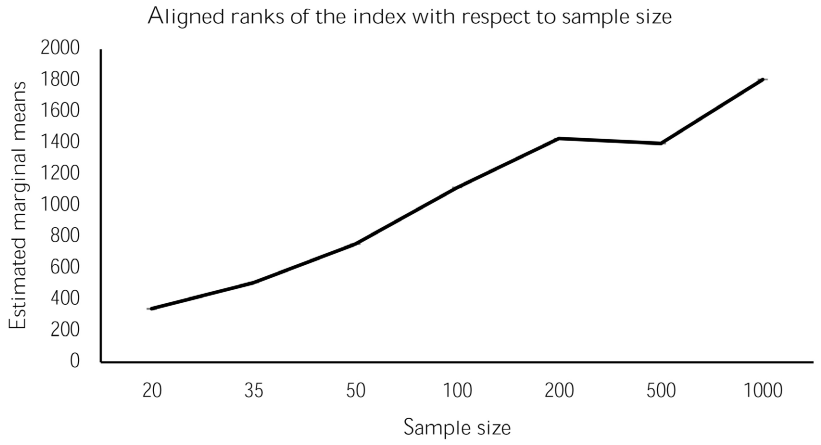

Figure 16. Plot of means conditional on the Sample size of the ranks (aligned with respect to the Sample size) for accuracy index.

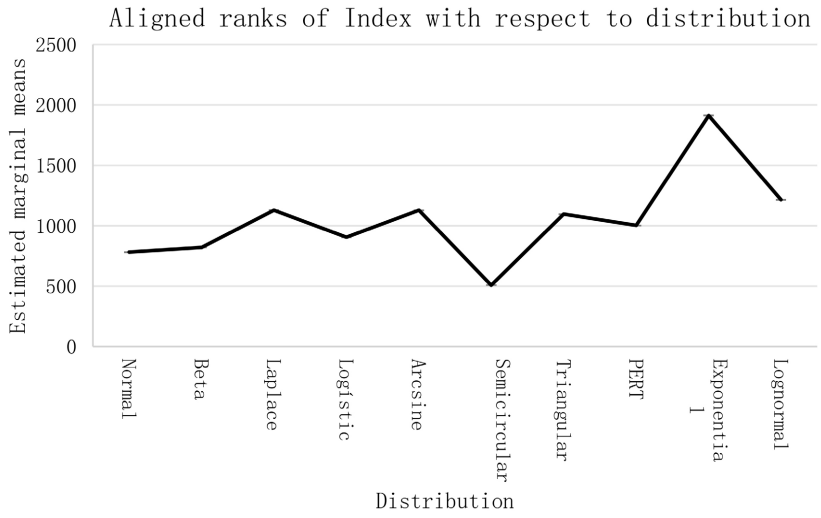

Figure 17. Plot of means conditional on the distribution type of the ranks (aligned with respect to the Distribution type) for accuracy index.

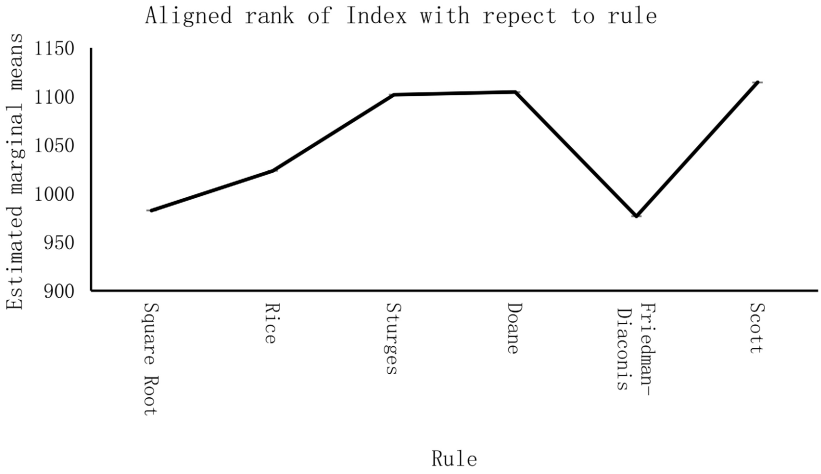

Figure 18. Plot of means conditional on the Rule of the ranks (aligned with respect to the Rule) for accuracy index.

The effect of the interaction between Distribution type and Rule is primarily attributed to the arcsine distribution. The highest mean ranks (aligned with respect to this interaction) for Accuracy Index appear with the square root, Rice, and Doane rules, and the lowest with the Friedman-Diaconis and Scott rules. The triangular distribution can also be highlighted, with an inverse pattern. Its highest mean ranks appear with the Friedman-Diaconis and Scott rules, and the lowest with the other three rules. The normal distribution performs best with Scott's rule and worst with the square root rule. The semicircular and PERT distributions favor the Friedman-Diaconis rule (Figure 19).

The effect of the interaction between Distribution type and Sample size mainly impacts the exponential and semicircular distributions. The former is better recognized with small sample sizes, and the latter with large sample sizes (Figure 20).

Regarding the interaction between Rule and Sample size, the Scott and Friedman-Diaconis rules benefit the least from the increase in sample size, while the Doane and Rice rules benefit the most. With a sample size of 20, the Friedman-Diaconis rule stands out, followed by the square root rule. With a sample size of 35, the square root rule excels. Scott's rule performs well with medium sample sizes of 50, 100, and 200 data points, while Doane's rule excels with large sample sizes of 500 and 1000 data points. Starting at a size of 200, the pattern is clearly ascending for the Doane and Rice rules, and with some ambiguity at the end for the Sturges rule. In contrast, it is clearly descending for the Scott and Freedman-Diaconis rules, and with some ambiguity at the end for the square root rule (Figure 21).

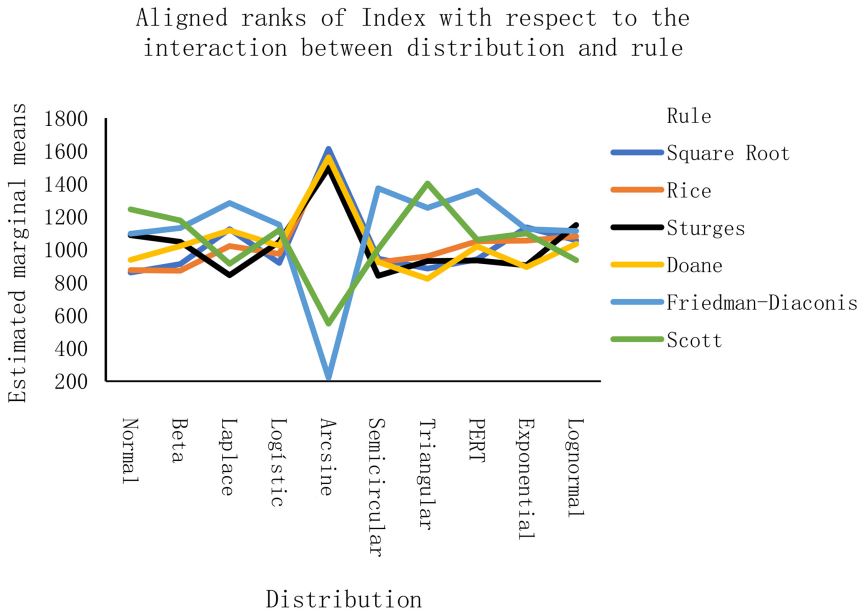

Figure 19. Plot of means conditional on the distribution type and rule of the ranks (aligned with respect to the interaction between the distribution type and rule) for accuracy index.

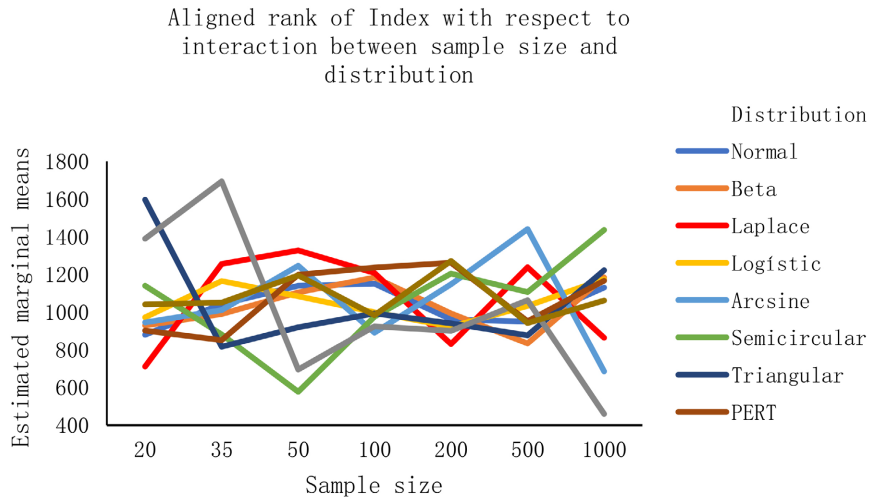

Figure 20. Plot of means conditional on the sample size and distribution type of the ranks (aligned with respect to the interaction between the Sample size and Distribution type) for accuracy index.

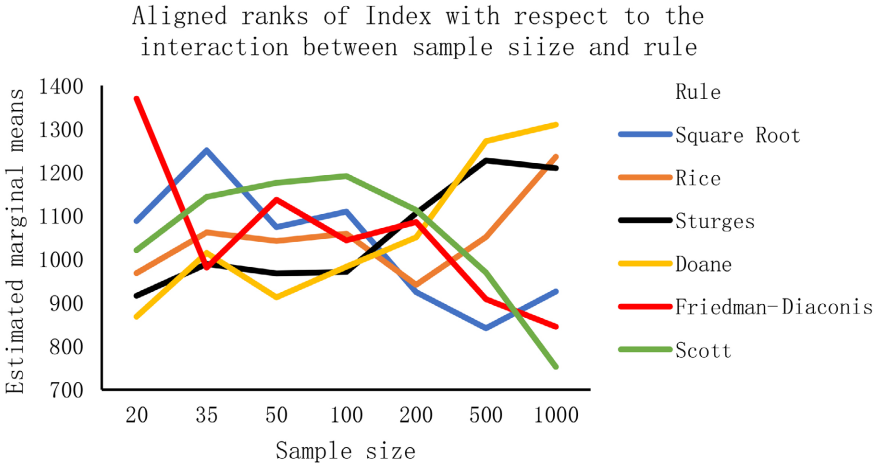

Figure 21. Plot of means conditional on the sample size and rule of the ranks (aligned with respect to the interaction between the Sample size and Rule) for accuracy index.

6. Conclusions

The AD statistic is strongly influenced by sample size, exhibiting a linear relationship with a very high strength of association. To a much lesser extent, it is affected by distribution type and the rule used to determine the number and amplitude of class intervals, with these two factors having small effect sizes. The arcsine distribution, characterized as a bimodal platykurtic distribution, generates the maximum discrepancy, while the exponential distribution achieves the minimum discrepancy. Platykurtic distributions exhibit more discrepancy than leptokurtic distributions. The square root, Rice, and Doane rules, with means equivalent to each other, have significantly less discrepancy than the Scott, Friedman-Diaconis, and Sturges rules. Consequently, Rice's rule, which can be based on the Scott and Friedman-Diaconis rules, generates less discrepancy be-

tween the empirical and theoretical histograms than these two rules. The interaction between rule and distribution type, as well as the triple interaction, are not significant. The second-order interactions of sample size with the rule and the distribution type are significant and small in size. The increase in sample size favors platykurtic distributions more, as well as the Scott and Friedman rules, that is, the conditions that generate the most discrepancy.

The Recognition Level is influenced by the sample size, with a large effect size. The relationship between sample size and Recognition Level is linear, demonstrating a strong association, although significantly lower than the Average Discrepancy. The type of distribution has a significant and large effect, with the least recognition given to the platykurtic semicircular distribution and the highest recognition to the leptokurtic exponential distribution. Rule type also has a significant, but small, effect size. Scott and Sturges rules have the highest recognition level, while the square root and Friedman-Diaconis rules have the least. Intermediate recognition levels are observed with Doane and Rice rules.

The increase in sample size primarily favors the recognition of the semicircular distribution. With small samples, the exponential and triangular distributions are the best recognized. The increase in size mainly favors the Doane and Rice rules and harms the Scott, Friedman-Diaconis, and square root rules. In the interaction between rule and distribution, it should be noted that the arcsine distribution, being a bimodal distribution, is poorly recognized by the Scott and Friedman rules, whereas the recognition of the triangular distribution is favored by these two rules. Recognition of Rice's rule is independent of the distribution.

Consistent with the results for Average Discrepancy and Recognition Level, the sample size has a significant and large effect on Accuracy Index. The relationship between sample size and the Accuracy Index is a direct linear one with a very large strength of association. The distribution type also has a significant and large effect, with the lowest accuracy occurring with the symmetrical and platykurtic semicircular distribution, and the highest accuracy with the positive asymmetric and leptokurtic exponential distribution. However, as with the Recognition Level, no greater accuracy is observed with leptokurtic distributions than with platykurtic ones.

The effect of the Rule on Accuracy Index is significant, but small. The Accuracy Index achieves its highest accuracy with the Scott, Doane, and Sturges rules, and its lowest with the Friedman-Diaconis and square root rules, with Rice's rule falling in between. Second-order interactions are significant. The increase in sample size favors the distribution where the Index has less accuracy, the semicircular one, and the Doane, Sturges, and Rice rules. The distribution where the Index shows more accuracy, which is the exponential, and the rules of Scott, Friedman-Diaconis, and square root, benefit less from the increase in population size. The rules of Scott and Friedman-Diaconis perform poorly with the arcsine distribution but show the greatest accuracy with the triangular one.

Rice's rule improves with increasing sample size. It performs better than the

Friedman-Diaconis rule, especially in relation to the arcsine distribution, but not as well as Scott's rule, except with this same distribution. In terms of the Accuracy Index, its profile resembles that of the square root rule across the 10 distributions and is akin to Doane's rule concerning sample size, differing from the profiles of the Scott and Friedman-Diaconis rules, which resemble each other. Among the seven rules, Scott's stands out, except with the arcsine distribution. In this case, the square root and Rice's rules are the better options. Consistent with other studies [19], the square root rule exhibits the lowest accuracy, except with the arcsine distribution, and is the one that benefits the least from an increase in the sample.

As limitations of the study, it should be noted that rules based on the minimization of arguments, such as the rules of Rudemo [16], Shimazaki and Shinomoto [17], Liu, Hussain, Tan, and Dash [18], or Knuth [21], were not taken into consideration. This decision was made due to their complexity in programming and their absence in the statistical packages currently in use. Additionally, the rule based on a homogeneous density but a heterogeneous width of k bins [15] [45] was also excluded, since it is limited to goodness-of-fit tests. Very large sample sizes, such as 2000, 5000, 10,000, or more data points, were not included either. Nevertheless, the scope of the present study aligns with common data analyses in research in psychology and related fields. In these domains, rules determining the number of bins and sample sizes, such as the aforementioned, are not commonly encountered.

The accuracy of the empirical histogram in reproducing the shape of the distribution was assessed through average discrepancy (between the empirical and expected histogram), recognition level (of the theoretical histogram), and an accuracy index (a combination of the two previous variables). However, there are other measures, such as integrated mean square error or the Kullback-Leibler divergence [46], which were more related to empirical rules that are not considered due to their computational complexity.

Acknowledgements

The author expresses gratitude 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] Lane, D.M. (2015) Guidelines for Making Graphs Easy to Perceive, Easy to Understand, and Information Rich. In: McCrudden, M.T., Schraw, G. and Buckendahl, C., Eds., *Use of Visual Displays in Research and Testing: Coding, Interpreting, and Reporting Data*, Information Age Publishing, Charlotte, 47-81.

[2] Lane, D. (2015) Histograms. Rice University, Houston.

https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Lane)/02%3A_Graphing_Distributions/2.04%3A_Histograms

[3] Tellechea-Robles, L.E., Salazar-Ceseña, M., Bullock, S.H., Cadena-Nava, R.D. and Méndez-Alonzo, R. (2020) Is Leaf Water-Repellency and Cuticle Roughness Linked to Flooding Regimes in Plants of Coastal Wetlands? *Wetlands*, **40**, 515-525. https://doi.org/10.1007/s13157-019-01190-7

[4] Sahann, R., Müller, T. and Schmidt, J. (2021) Histogram Binning Revisited with a Focus on Human Perception. *Proceedings of the* 2021 *IEEE Visualization Conference* (*VIS*), New Orleans, 24-29 October 2021, 66-70. https://doi.org/10.1109/VIS49827.2021.9623301

[5] Pearson, K. (1892) The Grammar of Science. Walter Scott Publishing Co., London. https://doi.org/10.1037/12962-000

[6] Sturges, H.A. (1926) The Choice of a Class Interval. *Journal of the American Statistical Association*, **21**, 65-66. https://doi.org/10.1080/01621459.1926.10502161

[7] Doane, D.P. (1976) Aesthetic Frequency Classification. *The American Statistician*, **30**, 181-183. https://doi.org/10.1080/00031305.1976.10479172

[8] Scott, D.W. (1979) On Optimal and Data-Based Histograms. *Biometrika*, **66**, 605-610. https://doi.org/10.1093/biomet/66.3.605

[9] Freedman, D. and Diaconis, P. (1981) On the Histogram as a Density Estimator: L_2 Theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete/Journal for Probability Theory and Related Fields*, **57**, 453-476. https://doi.org/10.1007/BF01025868

[10] Guerry, A.M. (1833) Essai sur la Statistique Morale de la France. Crochard, Paris.

[11] Nightingale, F. (1859) A Contribution to the Sanitary History of the British Army During the Late War with Russia, John W. Parker and Son, England.

[12] Rufilanchas, D. (2017) On the Origin of Karl Pearson's Term "Histogram". *Estadística Española*, **59**, 29-35.

[13] Magnello, M.E. (1996) Karl Pearson's Gresham Lectures: W. F. R. Weldon, Speciation and the Origins of Pearsonian Statistics. *The British Journal for the History of Science*, **29**, 43-63. https://doi.org/10.1017/S0007087400033859

[14] Ioannidis, Y. (2003) The History of Histograms (Abridged). *Proceedings of* 2003 *VLDB Conference*, Berlin, 9-12 September 2003, 19-30. https://doi.org/10.1016/B978-012722442-8/50011-2

[15] Moore, D.S. (1986) Tests of Chi-Squared Type. In: D'Agostino, R.B. and Stephens, M.A., Eds., *Goodness-of-fit- Techniques*, Marcel Dekker, New York, 63-95. https://doi.org/10.1201/9780203753064-3

[16] Rudemo, M. (1982) Empirical Choice of Histograms and Kernel Density Estimators. *Scandinavian Journal of Statistics*, **9**, 65-78. https://www.jstor.org/stable/4615859

[17] Shimazaki, H. and Shinomoto, S. (2007) A Method for Selecting the Bin Size of a Time Histogram. *Neural Computation*, **19**, 1503-1527. https://doi.org/10.1162/neco.2007.19.6.1503

[18] Liu, H., Hussain, F., Tan, C.L. and Dash, M. (2002) Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery*, **6**, 393-423. https://doi.org/10.1023/A:1016304305535

[19] Li, H., Munk, A., Sieling, H. and Walther, G. (2020) The Essential Histogram. *Biometrika*, **107**, 347-364. https://doi.org/10.1093/biomet/asz081

[20] Mohammed, M.B., Subhi, M.J. and Jamsari, A.A.W. (2022) New Approaches in

Frequency Table Construction for Continuous Symmetrical Data. *MATEMATIKA*: *Malaysian Journal of Industrial and Applied Mathematics*, **38**, 181-193. https://matematika.utm.my/index.php/matematika/article/view/1415

[21] Knuth, K.H. (2019) Optimal Data-Based Binning for Histograms and Histogram-Based Probability Density Models. *Digital Signal Processing*, **95**, Article 102581. https://doi.org/10.1016/j.dsp.2019.102581

[22] Stuart, A. and Ord, K. (2010) Kendall's Advanced Theory of Statistics: Volume 1: Distribution Theory. 6th Edition, John Wiley and Sons, New York.

[23] RDocumentation (2020) Quantile: Sample Quantiles. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/quantile

[24] George, D. and Mallery, P. (2022) IBM SPSS Statistics 27 Step by Step: A Simple Guide and Reference. 17th Edition, Routledge, New York.

[25] Zaiontz, C. (2018) Ranking Functions in Excel. Real Statistics Using Excel. https://real-statistics.com/descriptive-statistics/ranking-function-excel/

[26] Hyndman, R.J. and Fan, Y. (1996) Sample Quantiles in Statistical Packages. *The American Statistician*, **50**, 361-365. https://doi.org/10.2307/2684934

[27] Zaiontz, C. (2021) Harrell-Davis Quantiles. Real Statistics Using Excel. https://real-statistics.com/descriptive-statistics/ranking-function-excel/harrell-davis-quantiles/

[28] Schlunk, S. and Byram, B. (2022) Breaking and Fixing gCNR and Histogram Matching. *Proceedings of the* 2022 *IEEE International Ultrasonics Symposium* (*IUS*), Venice, 10-13 October 2022, 1-4. https://doi.org/10.1109/IUS54386.2022.9958858

[29] Paulauskas, N. and Baskys, A. (2019) Application of Histogram-Based Outlier Scores to Detect Computer Network Anomalies. *Electronics*, **8**, Article 1251. https://doi.org/10.3390/electronics8111251

[30] Deka, K., Shah, Z.A., Misra, R. and Ahmed, G.A. (2022) A Study of The Effects of Histogram Binning on the Accuracy of Finding Flux Distribution of X-Ray Binaries. *Materials Today*: *Proceedings*, **65**, 2862-2864. https://doi.org/10.1016/j.matpr.2022.06.279

[31] Ganesh, E.N. and Vistas, D. (2022) Image Registration of Medical Images Using Mutual Information Algorithm and Histogram Methods. 2*nd National Conference on Biomedical Engineering*, *National Institute of Technology*, Rourkela, 1 January 2022, 1-8. https://scholar.archive.org/

[32] Hyndman, H.J. (1995) The Problem with Sturges' Rule for Constructing Histograms. https://robjhyndman.com/papers/sturges.pdf

[33] Fulp, H. and Louise, M. (2021) Dynamic Reduction of Scientific Data through Spatiotemporal Properties. Thesis, Clemson University, Clemson. https://tigerprints.clemson.edu/all_theses/3656

[34] Pearson, K. (1895) Contributions to the Mathematical Theory of Evolution II: Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London*, *Series A*, **186**, 343-414. https://doi.org/10.1098/rsta.1895.0010

[35] Pearson, E.S. (1931) Note on Tests for Normality. *Biometrika*, **22**, 423-424. https://doi.org/10.1093/biomet/22.3-4.423

[36] Royston, P. (1993) A Toolkit for Testing for Non-Normality in Complete and Censured Samples. *Journal of the Royal Statistical Society*, *Series D* (*The Statistician*), **42**, 37-43. https://doi.org/10.2307/2348109

[37] D'Agostino, R.B., Berlanger, A. and D'Agostino, R.B. (1990) A Suggestion for Using Powerful and Informative Test of Normality. *The American Statistician*, **44**, 316-321.

https://doi.org/10.2307/2684359

[38] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. (2011) The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, May 2011, 143-146. https://doi.org/10.1145/1978942.1978963

[39] Meier, U. (2006) A Note on the Power of Fisher's Least Significant Difference Procedure. *Pharmaceutical Statistics*: *The Journal of Applied Statistics in the Pharmaceutical Industry*, **5**, 253-263. https://doi.org/10.1002/pst.210

[40] Holm, S. (1979) A Simple Sequentially Reject Procedure. *Scandinavian Journal of Statistics*, **6**, 65-70.

[41] Cohen, J. (1992) A Power Primer. *Psychological Bulletin*, **112**, 155-159. https://doi.org/10.1037/0033-2909.112.1.155

[42] Ben-Shachar, M.S., Lüdecke, D. and Makowski, D. (2020) Effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, **5**, Article 2815. https://doi.org/10.21105/joss.02815

[43] Rosner, B. and Glynn, R.J. (2006) Interval Estimation for Rank Correlation Coefficients Based on the Probit Transformation with Extension to Measurement Error Correction of Correlated Ranked Data. *Statistics in Medicine*, **26**, 633-646. https://doi.org/10.1002/sim.2547

[44] Meng, X.-L., Rosenthal, R. and Rubin, D.B. (1992) Comparing Correlated Correlation Coefficients. *Psychological Bulletin*, **111**, 172-175. https://doi.org/10.1037/0033-2909.111.1.172

[45] Sulewski, P. (2021) Equal-Bin-Width Histogram versus Equal-Bin-Count Histogram. *Journal of Applied Statistics*, **48**, 2092-2111. https://doi.org/10.1080/02664763.2020.1784853

[46] Kim, T., Oh, J., Kim, N., Cho, S. and Yun, S.Y. (2021) Comparing Kullback-Leibler Divergence and Mean Squared Error Loss in Knowledge Distillation. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 2628-2635. https://doi.org/10.48550/arXiv.2105.08919