standards

Article

Composition of Probabilistic Preferences in Multicriteria Problems with Variables Measured in Likert Scales and Fitted by Empirical Distributions

Luiz Octávio Gavião [1], Annibal Parracho Sant'Anna [2,*], Gilson Brito Alves Lima [2] and Pauli Adriano de Almada Garcia [3]

[1] Science & Technology Division, Brazilian War College, Rio de Janeiro 22291-090, Brazil; luiz.gaviao67@gmail.com
[2] Industrial Engineering Department, Fluminense Federal University, Niterói 24220-240, Brazil; glima@id.uff.br
[3] Administration Department, Fluminense Federal University, Volta Redonda 27213-145, Brazil; pauliadriano@gmail.com
* Correspondence: annibal.parracho@gmail.com

Abstract: The aim of this article is to demonstrate the advantages of the Composition of Probabilistic Preferences method in multicriteria problems with data from Likert scales. Multicriteria decision aids require a database as a decision matrix, in which two or more alternatives are evaluated according to two or more variables selected as decision criteria. Several problems of this nature use measures by Likert scales. Depending on the method, parameters from these data (e.g., means, modes or medians) are required for calculations. This parameterization of data in ordinal scales has fueled controversy for decades between authors who favor mathematical/statistical rigor and argue against the procedure, stating that ordinal scales should not be parameterized, and scientists from other areas who have shown gains from the process that compensates for this relaxation. The Composition of Probabilistic Preferences can allay the protests raised and obtain more accurate results than descriptive statistics or parametric models can bring. The proposed algorithm in R-code involves the use of probabilities with empirical distributions and fitting histograms of data measured by Likert scales. Two case studies with simulated datasets having peculiar characteristics and a real case illustrate the advantages of the Composition of Probabilistic Preferences.

Keywords: probabilistic preferences; CPP; Likert scales; empirical distributions

check for updates

Citation: Gavião, L.O.; Sant'Anna, A.P.; Lima, G.B.A.; Garcia, P.A.d.A. Composition of Probabilistic Preferences in Multicriteria Problems with Variables Measured in Likert Scales and Fitted by Empirical Distributions. *Standards* **2023**, *3*, 268–282. https://doi.org/10.3390/standards3030020

Academic Editor: Michael Glykas

Received: 21 May 2023
Revised: 30 June 2023
Accepted: 7 July 2023
Published: 17 July 2023

1. Introduction

In this article, we propose the use of the Composition of Probabilistic Preferences (CPP) in multicriteria aid problems that measures their variables on Likert scales. In general, Likert scale responses are summarized as means, used as input in some decision support models. Adapting the spectrum of Likert scale responses to a probability distribution, thereby averting the controversies surrounding the use of descriptive statistics derived from Likert scales, is the innovation of the current study. This Likert scale approach in multicriteria decision aid methods has not been found in the literature so far. The contribution of the research lies in the ability of CPP to mitigate the controversies about the use of Likert scales for decision aid problems and to obtain more accurate results than descriptive statistics or parametric models can bring.

The authors' motivation was to demonstrate how CPP is tailored to solve these controversies, based on its own features. CPP is the most suitable method for this approach, as only CPP uses probability distributions as inputs in its algorithm [1,2]. A description of Likert scales, the controversies in the literature, and the advantages of approaching the problem with CPP are presented below.

Likert scaling was introduced by Rensis Likert in 1932 [3] and is the most widely used psychometric scale in survey research, where respondents indicate their levels of agreement with a statement and label them on a scale [4,5]. For example, a five-point Likert scale could be labeled as "1" (totally disagree), "2" (disagree), "3" (neutral), "4" (agree), and "5" (strongly agree). Depending on what is being measured, the scale labels may be worded differently. Likert also considered the questionnaire responses through normal distributions, resorting to the use of means and standard deviation, for example, to draw conclusions [3].

In multicriteria decision aid problems, it is necessary to compose a decision matrix, in which two or more alternatives are evaluated according to two or more variables selected as decision criteria [1]. To solve the problem, those variables must be measured on some scale so that a calculation algorithm can generate results and reach the objective of the research. Several studies on multicriteria decision aid problems use variables measured by Likert scales [6–9].

From a statistical point of view, a Likert scale is considered an ordinal scale, as described in Table 1 [10]. Therefore, the use of parametric models (like the normal distribution, which depends on the mean and standard deviation of the data) with variables measured on ordinal scales is an error [11–14]. This common practice in scientific studies was even described as the first mathematical "sin" on a list of seven [15]. A brief discussion about the controversies of this use of Likert scales is presented here.

Table 1. Types of scales [10].

Scale	Basic Empirical Operations	Permissible Statistics
Nominal	Determination of equality	Number of cases Mode Contingency correlation
Ordinal	Determination of greater or less	Median Percentiles
Interval	Determination of equality of intervals or differences	Mean Standard deviation Product-moment correlation
Ratio	Determination of equality of ratios	Coefficient of variation

The research was guided by three questions: (1) How do the main controversies about the use of Likert scales to support decisions evolve? (2) Which methodological changes in CPP can enhance the use of data based on Likert scales? (3) What are the practical advantages of using empirical distributions in CPP?

2. Main Statistical Controversies

In general, the main criticism is the lack of mathematical/statistical rigor in relation to the measurement of variables on an ordinal scale. In this case, it is inappropriate to calculate averages, standard deviations, and other measures of central tendency to draw conclusions that have little or no meaning in the real world. For example, if a certain surgical procedure is rated on a four-point psychometric scale (e.g., poor, fair, good, or excellent), it would not make sense to average the responses to the procedure, yielding a "fair and medium" result [15].

However, there are several authors who defend the use of descriptive statistics and parametric models in structured problems with Likert scales. Some argue that the use of scales with more than five points can better approximate the results to a normal distribution [16,17], or favor parametric models over nonparametric ones [18]. Other authors have mentioned the amount of research that "violates" the principles of statistics and obtained satisfactory results [19], the constraints on the use of parametric models with Likert scales [20], or the context in which such restrictions were proposed for the use of these scales [21], criticizing the relaxation of mathematical rigor to the detriment of practical benefits.

The parameterization of data measured on ordinal scales has fueled controversy for decades between authors who favor mathematical/statistical rigor and argue against the procedure, considering that ordinal scales should not be parameterized, and academics from other areas who have reported gains in their studies that compensate for the relaxation of this rigor. We selected some studies defending the use of more flexible measures of central tendency in data generated by Likert scales and others against the relaxation of mathematical/statistical rigor in these problems to provide a better understanding of the arguments.

One of the most interesting papers on this controversy is from Thomas [21]. The author describes these discussions in the early 1900s, when physics was the model science and measurement followed mathematical principles to draw empirical conclusions about psychological data. Thomas [21] argued that psychology should treat measurement as the recording of empirical facts, which may not meet the mathematical properties of real numbers. There was then some error of interpretation, adaptation, or an excess of rigor in the use of the four measurement scales (nominal, ordinal, interval, and ratio), proposed by Stevens [10], according to Table 1.

Thomas [21] pointed out that the article of Stevens [10] helped create a bridge to the use of mathematics to draw empirical conclusions about data without the properties of real numbers. However, this study was a product of its time and urged limitations on the mathematical operations for each scale, without explicitly anchoring those limitations in mathematics. Thomas [21] continued this criticism, noting that Stevens's scales are flawed and encouraging the production of academic works that do not conform to the epistemology of science or the logic of mathematics. Other authors have focused their research on finding solutions that allow the use of parameters based on Likert scales. Wu and Leung [16] tried to adapt ordinal scales as interval scales, increasing the number of points on the scale, preferably to eleven, to facilitate the adjustment of data to normal and other parametric distributions. Along the same lines, Awang et al. [17] explored a parametric model of structural equations to suggest the use of ten-point Likert scales instead of the traditional five-point scales, to obtain more consistent results. Harpe [20] reviewed the literature, and based on empirical evidence, concluded that parametric analytical approaches are acceptable as long as certain criteria are met. The author also explored histogram densities, similar to the proposal of this article, but with adjustments to normal distributions instead of nonparametric ones. Mircioiu and Atkinson [22] analyzed ordinal data on Likert scales with high response rates, demonstrating that analysis using nonparametric methods causes a loss of information. The addition of parametric methods, graphical analysis, subset analysis, and data transformation lead to deeper analysis and better conclusions.

The authors who defend greater mathematical/statistical rigor condemn the use of models and parametric measures with data collected from Likert scales, regardless of the number of points on the scale. Pornel and Saldaña [23] discussed the characteristics and proper use of a Likert scale, examining 53 theses and dissertations in the Philippines. They identified four common misuses of the Likert scale, namely: an unjustified length of the scale; asymmetrical verbal anchoring; uneven spacing in the verbal anchor; and unjustified interpretation of the mean. Jamieson [24] emphasized the common practice of assuming that the Likert scale is an interval scale, condemning the calculation of parametric measures and exemplifying the means and standard deviations, since these are typical of ordinal scales. Allen and Seaman [25] followed the same line. Sullivan and Artino [26] analyzed research in the medical field and made the use of parametric tests more flexible, although they also criticized the use of means and standard deviations.

As can be seen, the subject is still controversial in the scientific literature [27,28]. Two partial conclusions can be drawn from the studies for and against the full use of statistical tools with data measured on Likert scales: (1) the nonparametric approach avoids difficulties present in parametric procedures; and (2) the ranking of alternatives is consistent with the possible answers of ordinal scales, as defined by Stevens [10].

The use of CPP fulfills these needs, since it is flexible regarding the choice of parametric or nonparametric models and can be used to define the probability of an alternative being better or worse than others, which corresponds to the empirical possibilities described in the ordinal scale of Stevens [10]. We present two cases to illustrate these characteristics numerically in Section 5.

There are basically three advantages in the algorithm proposed here: (1) it is a nonparametric procedure, maintaining adherence to statistics with regard to the treatment of data on ordinal scales, without resorting to parameters to solve the problem (e.g., mean, deviation-default); (2) it dispenses with assigning numerical values to the Likert scale, since the frequency of responses to each item of the scale is sufficient; and (3) it explores an important property of the CPP, which probabilistically models responses to consider the uncertainty intrinsic to human opinion or judgment due to prejudices, doubts, and biases, among other aspects that indicate a nondeterministic approach to the problem. The main disadvantage of this approach with CPP relates to the complexity of the calculations, compared to the simple use of descriptive statistics, available in commercial spreadsheet-like applications. However, this disadvantage was mitigated by publishing the proposed software as open source on the Zenodo.org platform [29] (and making it available in Appendix A here).

3. The Basics of CPP

The proposed model is based on CPP, originally developed by Sant'Anna and Sant'Anna [30] and later expanded by Sant'Anna [1]. This method incorporates the probabilistic nature of preference assessment into the multicriteria decision aid problem. The probabilistic characteristic can arise because of inaccuracies caused by subjective factors, leading decision makers to attribute different meanings to the same attributes of alternatives in different circumstances, or because of measurement errors that affect the evaluations of such attributes.

An initial and critical step in CPP is the transformation of the numerical vector, containing the evaluations of the various alternatives according to each criterion, into a vector containing preference probabilities. This transformation of exact values into probabilities is illustrated in Figure 1, where different distributions emulate the preferences of an expert. The exact value that corresponds to an expert's assessment becomes the mode of a distribution of preferences, which varies between the extremes of the assessments by the criterion. To illustrate this procedure, the exact value "4" attributed to Alternative A and the range from "1" to "5" of Criterion 1 are converted into three probability distributions: a Triangular distribution of parameters (1, 4, 5), which indicate the minimum value "1", the mode "4", and the maximum value "5"; a Beta PERT distribution with those same parameters plus a fourth one, called shape, that emulates the variation in values around the mode; and a Normal distribution, of parameters (4, 2), where "4" is the mean and "2" simulates the standard deviation of evaluations on the same criterion. Any type of probability distribution can be used with CPP, both parametric and non-parametric, giving the method a significant versatility.

After randomizing the variables, CPP makes a relative comparison among the alternatives' performances in each criterion. In this second stage, it is possible to verify, for example, to what extent each alternative could maximize or minimize its random performance to all the others. The calculation, for each criterion, is performed by integrating a function that corresponds to the product of the probability density of the alternative considered and the cumulative function of the other alternatives, as described by Equations (1) and (2). In the calculus of probabilities of the maximization and minimization—that is, the i-th alternative is superior ($PMax_{ij}$) and inferior ($PMin_{ij}$) to all the others according to the j-th criterion—the evaluation of each alternative is represented by a random vector X; the functions of the i-th alternative are indexed by "i" and the functions of all the others by

"−i". The Cumulative Distribution Function (CDF) is represented by F_X, the Probability Density Function (PDF) by f_X, and the domain of the random variable X is represented by D_X [31].

$$PMax_{ij} = \int\limits_{D_{X_i}} \left[\prod F_{X_{-i}}(x_{-i})\right] f_{X_i}(x_i)dx_i \tag{1}$$

$$PMin_{ij} = \int\limits_{D_{X_i}} \left[\prod (1 - F_{X_{-i}}(x_{-i}))\right] f_{X_i}(x_i)dx_i \tag{2}$$

In the last stage, the probabilities of each alternative being the best or worst in each criterion are combined in different ways, providing different points of view for decision-making. For instance, PMax and PMin can be composed by axis (progressive–conservative/optimist–pessimist) [32], by principles of concentration/dilution [33], by Choquet capacities [34], by the Gini index [35], by classes [2], and associated with other methods [36]. Automation of the complete CPP procedure is available with open access in the "CPP" package [37]. This research focuses on the first stage of CPP, as detailed in the next section.

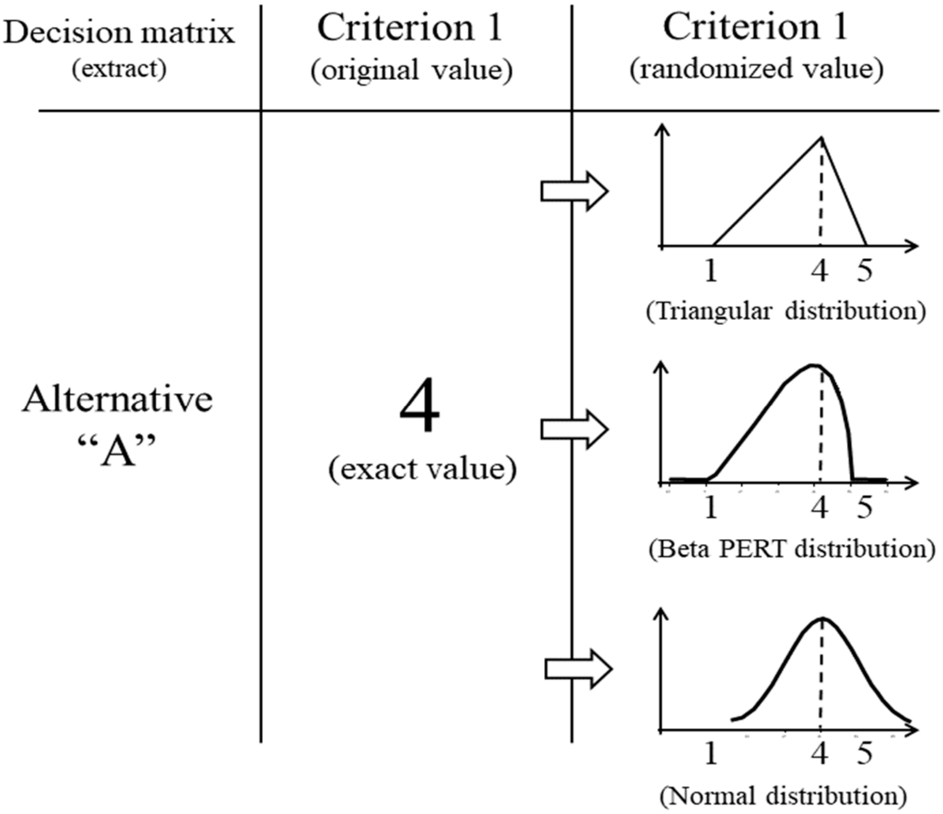

Figure 1. CPP—first stage.

4. Materials and Methods

The new approach, in the lower portion of Figure 2, proposes a nonparametric procedure for the first stage of CPP, for cases in which the respondents' original assessments are based on a Likert scale. Instead of considering each respondent's preference for conversion into a parametric distribution, the histogram of the dataset is transformed into a nonparametric distribution. This conversion of the histogram into a nonparametric distribution is found in the literature [38–41], but its adaptation to CPP is unprecedented.

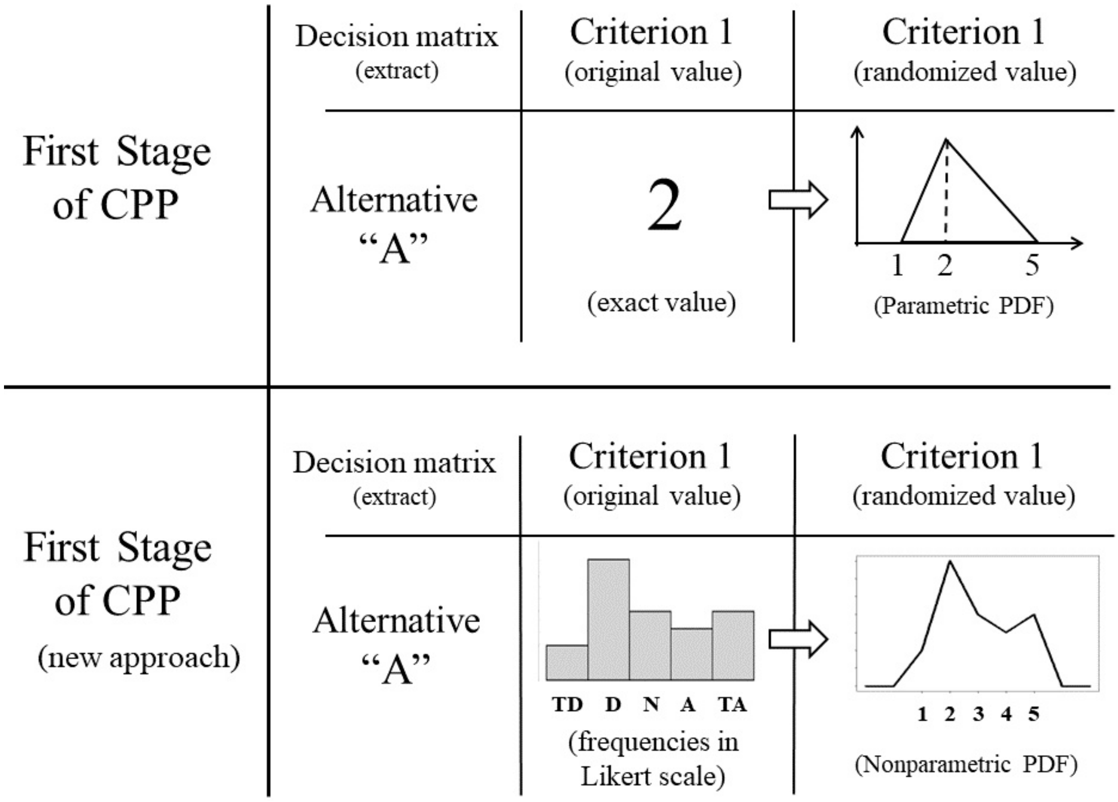

Figure 2. New approach of the first stage. Abbreviations: TD (totally disagree), D (disagree), N (neutral), A (agree), TA (totally agree).

In Figure 2, we see that the measure of descriptive statistics (e.g., means) used to represent a spectrum of responses in the Likert scale is discarded. The histogram of responses is fitted to a probability distribution, reflecting the full density of data, not just one measure.

This approach values CPP. The choice of nonparametric PDF avoids the controversy described in Section 2, as it does not use descriptive statistics to extract parameters from the dataset. In addition, the new model simplifies CPP calculations in problems with a high number of respondents. For example, hundreds of evaluations of an alternative in a criterion can be generalized in a single histogram, avoiding the same number of calculations or the need to resort to a measure of central tendency to represent them.

The nonparametric distribution chosen in this approach is an empirical distribution, in which each point on the Likert scale is associated with a probability, calculated from the responses obtained from questionnaires or interviews. The R software version 4.3.1 provides several applications for fitting data to probability functions. The empirical distributions fitted here are based on the "mc2d" package [42].

The two subsequent stages of CPP remain unchanged and are described in detail in several published articles and books, with a wide range of applications, including sports science [35], management systems [43,44], security and defense [45], public health and social assistance [46,47], risk management [48], and bioenergy processes [36].

Table 2 reports the pseudocode used to calculate the preference probabilities, referring to the second stage of the CPP, according to the proposed approach. The calculations were performed with the R software. The third stage of the CPP was not included in the algorithm, since the examples in this article refer to the evaluation of alternatives under only one criterion. This final aggregation of the third stage of CPP requires evaluations of two or more criteria, beyond the scope of this proposal.

Table 2. Pseudocode for Second Stage of the CPP.

Empirical Probabilities of Preference on Likert Scales
1. Description: ranking alternatives evaluated on a criterion
2. Variables
>values—vector with numerical sequence of Likert scale options
>freqs—Likert scale option frequency matrix:
-matrix rows: problem alternatives
-matrix columns: frequencies of Likert scale options
3. Commands
>open the R software console
>install the R software "mc2d" library
>load the database "values" and "freqs"
>run the "PMax.Emp.Likert" function, for "benefit" type criteria
>run the "PMin.Emp.Likert" function, for "cost" type criteria
>rank alternatives in the criteria
4. End

The code registered on the Zenodo.org platform offers two functions: "PMax.Emp.Likert" to calculate the joint probabilities of maximizing the alternatives and "PMin.Emp.Likert" for minimizing the alternatives. Both are required for certain CPP compositions [1]. However, additional care must be taken when applying them to problems of ranking alternatives evaluated on Likert scales, related to the type of evaluation criterion. The "benefit" criteria indicate that higher scores on the scale are better for the decision. In this case, evaluations with a value of "5" are more important than those with a value of "4", and so on. On the other hand, the "cost" criteria indicate that lower scores on the scale are better for the decision. Finally, it should be noted that the joint probabilities of maximizing and minimizing, considering three or more alternatives, are not complementary.

5. Applications

5.1. Dataset with the Same Means and Medians

We generated two datasets to illustrate the usefulness of the proposal in multicriteria decision aid problems with the alternatives measured by Likert scales. The first case contains ten alternatives, evaluated by 500 respondents, on a symmetrical five-point scale: "totally disagree", "disagree", "neutral", "agree", and "totally agree". The peculiar aspect of this case is that the responses of the ten alternatives ("A" to "J") have the same average and the same median for any scale of values assigned to the five points, as long as they are equidistant. For example, if the five points assume values from "1" to "5", the mean and median will be equal to "3"; if the points assume values (1, 3, 5, 7, and 9), the mean and median will be "5". We chose this setup to demonstrate the limitation of descriptive statistics in comparison with the proposed model. Equations (1)–(10) indicate the frequencies of responses for each alternative, in which the first represents the option "totally disagree" and the last "totally agree". For example, Alternative "A" received 50 "totally disagree", 175 "disagree" responses, and so on.

$$A = \{50, 175, 100, 75, 100\} \tag{3}$$

$$B = \{150, 75, 75, 25, 175\} \tag{4}$$

$$C = \{100, 100, 125, 50, 125\} \tag{5}$$

$$D = \{100, 100, 75, 150, 75\} \tag{6}$$

$$E = \{75, 50, 200, 150, 25\} \tag{7}$$

$$F = \{25, 175, 175, 25, 100\} \quad (8)$$

$$G = \{75, 0, 325, 50, 50\} \quad (9)$$

$$H = \{100, 100, 100, 100, 100\} \quad (10)$$

$$I = \{125, 75, 125, 25, 150\} \quad (11)$$

$$J = \{150, 0, 175, 50, 125\} \quad (12)$$

Figure 3 illustrates the histograms of Alternatives "A" and "B", in which each column indicates the proportion of choice of the respective scale point, in relation to the 500 responses. The curve to the right of each histogram is the empirical PDF, which is nonparametric in nature. An interesting aspect to visualize is that the PDFs are unique for each dataset, differentiating them even for the situation of equality of their means and medians. This kind of data "DNA" is sensitive to CPP, as its calculations depend on the PDF and its related cumulative function (CDF) to identify which alternative is likely to be superior or inferior to the others.

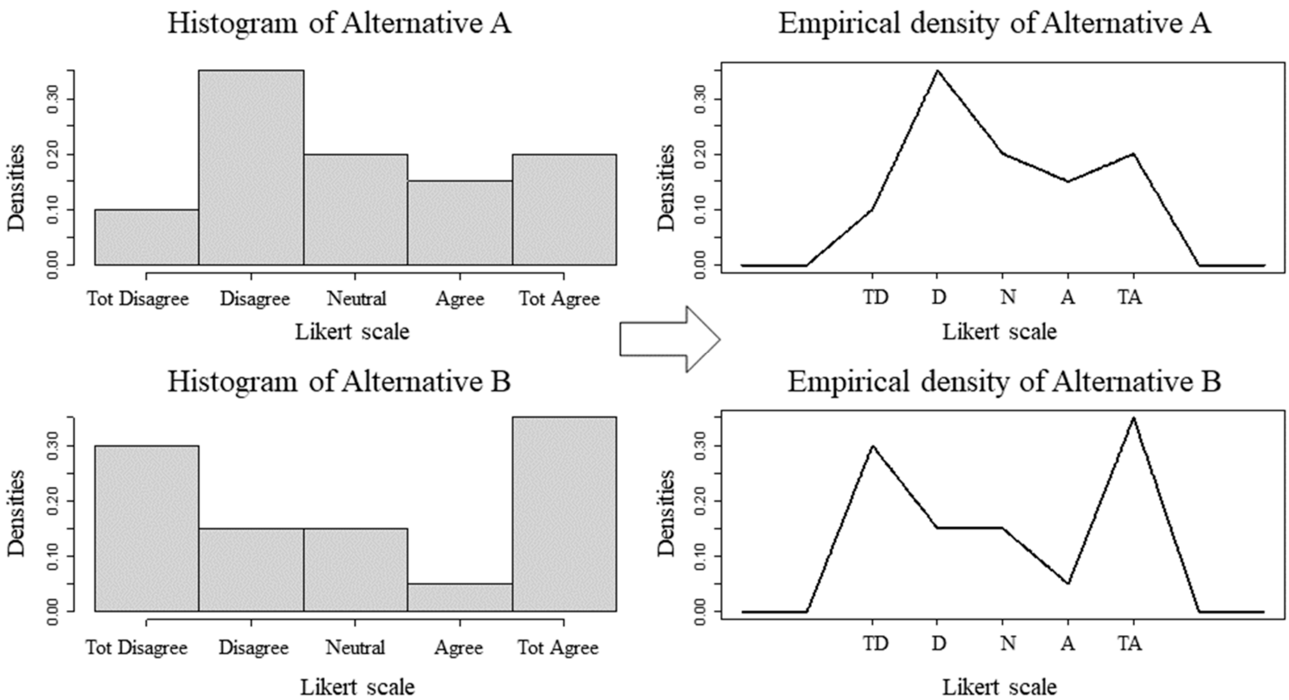

Figure 3. Histograms and densities of alternatives "A" and "B". Abbreviations: TD (totally disagree), D (disagree), N (neutral), A (agree), TA (totally agree).

The PDF of the histograms in Figure 3 was defined by the "dempiralC" function, available in the "mc2d" package [42], according to Equation (13), where "x" are the scale values and "p" denotes their probabilities of occurrence, based on their frequencies. The "*p*" values are normalized to give the distribution one unit of area. The functions "dempiralC" and "pempiricalC" (CDF) were used in the calculations of PMax and PMin, according to the R code in Appendix A.

$$f(x) = p_i + \left(\frac{x - x_i}{x_{i+1} - x_i}\right)(p_{i+1} - p_i),\ x_i < x < x_{i+1} \quad (13)$$

Once the probability distributions that fit the evaluations represented by the histograms have been defined, the next step of CPP refers to the calculation of the probabilities that the alternatives maximize or minimize their preferences, considering the type of criterion. Based on the code registered on the Zenodo.org platform, the functions "PMax.Emp.Likert" and "PMin.Emp.Likert" were applied to the ten alternatives, generating the results in Table 3. The highest frequency of extreme values of Alternative "B", with 150 "totally disagree" and 175 "totally agree" responses, were decisive for its priority, both to maximize and minimize its preference over the others.

Table 3. Results of the second stage of the CPP.

Alternative	PMax	PMin
A	0.08981332	0.09327040
B	0.15662032	0.16098827
C	0.10803823	0.10993096
D	0.10599313	0.10154487
E	0.06747378	0.06140966
F	0.07062573	0.07502867
G	0.05176373	0.04871787
H	0.10671008	0.10557457
I	0.12501971	0.12782754
J	0.11794092	0.11570593

5.2. CPP Sensitivity to Likert Scale Cardinality

The second illustration of the usefulness of the model proposed here explores two alternatives ("K" and "M"), with variants K* and M*, according to the frequencies indicated in Equations (12)–(15). Alternatives "K" and "M" were rated on a five-point scale and alternatives K* and M* on a nine-point scale. The extreme options remain with the indications "totally disagree" and "totally agree" and the intermediate ones as "neutral". The K* and M* frequencies repeat the values of the "K" and "M" evaluations, respectively, only representing greater rigor on the part of the evaluators by using the nine-point scale. Both "K" and "M" and K* and M* have the same medians, but different means and modes, which would be enough to define the priority for K or K*, if decision making prioritized these measures of descriptive statistics.

$$K = \{150, 50, 100, 25, 175\} \tag{14}$$

$$M = \{75, 50, 200, 150, 25\} \tag{15}$$

$$K^* = \{150, 50, 100, 25, 175, 0, 0, 0, 0\} \tag{16}$$

$$M^* = \{75, 50, 200, 150, 25, 0, 0, 0, 0\} \tag{17}$$

Figures 4 and 5 describe the histograms and empirical densities of alternatives K and M and their variants K* and M*, respectively.

In the literature review, we identified studies that sought to make the use of descriptive statistics more flexible in problems with Likert scales of higher cardinality, arguing that ten-point [17] or eleven-point [16] scales, for example, would be sufficient to relax the mathematical rigor/statistics of the scales of Stevens [10]. This application shows that the CPP is able to confirm this argument, but without resorting to statistical parameters for decision making.

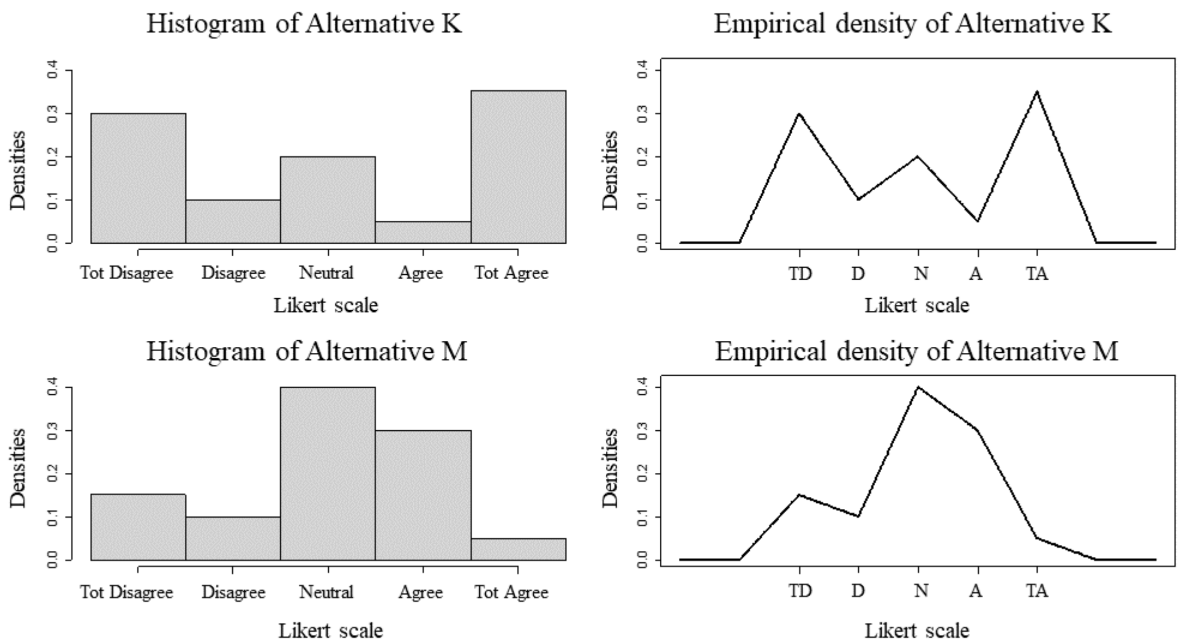

Figure 4. Histograms and densities of alternatives "K" and "M" (5-point scale). Abbreviations: TD (totally disagree), D (disagree), N (neutral), A (agree), TA (totally agree).

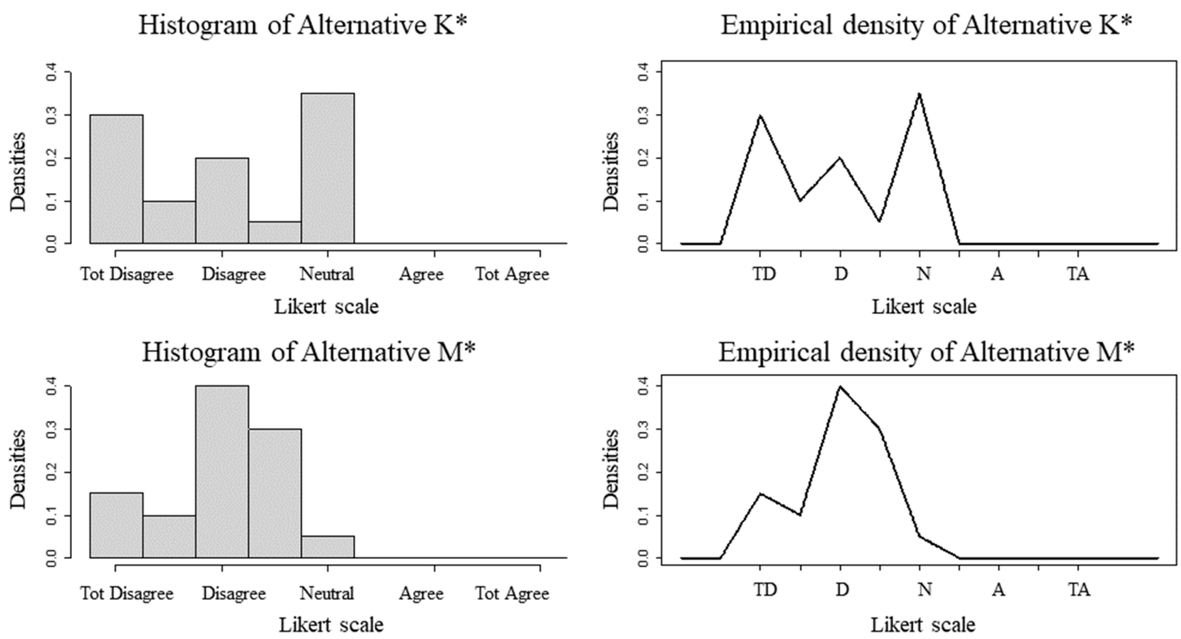

Figure 5. Histograms and densities of alternatives "K*" and "M*" (9-point scale). Abbreviations: TD (totally disagree), D (disagree), N (neutral), A (agree), TA (totally agree).

Table 4 indicates that the expansion of the Likert scale points was not enough to change the parameters of the variables (median, mean, and mode) in order to change the decision. However, the change in scale gives a new density to the data, even if the density profiles of K* and M* are similar (shifted to the left) to the respective ones of K and M. The CPP is able to capture this change in scale due to its nonlinear property, valuing the highest evaluations in the PMax calculations and valuing the smallest ones in the PMin calculations, as described in Gavião et al. [31]. Thus, in the new nine-point scale, the PMax of Alternative K* becomes higher than that of M*, confirming the order of priority indicated by means and modes, without eliciting the historical critiques of using descriptive statistics in ordinal scales.

Table 4. Results of the second stage of the CPP.

Scales	Alternative	Median	Mean	Mode	PMax	PMin
5 points	K	3	3.05	5	0.4667357	0.5332673
	M	3	3	3	0.5332673	0.4667357
9 points	K*	3	3.05	5	0.5637278	0.4362749
	M*	3	3	3	0.4362749	0.5637278

5.3. Unified Health System (SUS)

The national health system in Brazil is called the Unified Health System (SUS in the Portuguese abbreviation), ranging from basic care to complex procedures such as organ transplants. The SUS guarantees full, universal, and free access to the country's population. Its chain of health units includes federal university hospitals, which are reference centers of medium and high complexity for the SUS. In addition, these hospitals are important training centers for human resources in the health area and provide support to teaching, research, and extension of the federal institutions of higher education to which they are linked.

The network of federal university hospitals comprises 51 hospitals linked to 36 federal universities. Of these, 41 hospitals are linked to the Brazilian Hospital Service Company (Ebserh). This company was created by the federal government and is responsible for the management of federal university hospitals. Among the attributions assumed by Ebserh are the coordination and evaluation of execution of the hospitals' activities; technical support for the elaboration of management improvement instruments; and the distribution of resources to the hospitals. Several reports of satisfaction surveys carried out by Ebserh are made available for public access (https://www.gov.br/ebserh/pt-br/acesso-a-informacao/participacao-social/ouvidoria-geral/pesquisas-de-satisfacao, accessed on 11 April 2023).

In 2022, Ebserh published a panel of results with the satisfaction rates of different users of university hospitals, which included resident physicians in their units. The study consolidated the responses on a nine-point Likert scale, but drew conclusions based on calculating the averages of these responses. Thus, in this case study, we sought to apply the proposed method to reassess the final ranking of teaching hospitals, based on the opinion of resident physicians.

Table 5 presents the results of calculating the PMax for each hospital, compared with the final ranking tallied by Ebserh. Kendall's ordinal statistical correlation was applied to the two rankings, indicating a moderate relationship of 49.2%. Three hospitals had the same position in the satisfaction assessment, while 18, highlighted in this table, improved their order in the probabilistic calculations. The results illustrate the greater discriminant power of CPP for the final calculation. This is relevant, since these 18 hospitals could have priority in the distribution of public resources or in public educational policies, receiving new diagnostic equipment and medicines, among other benefits.

Table 5. Case study results.

Hospital	Likert Scale (% of Evaluations)					PMax	Rank PMax	Rank Ebserh	Ebserh Result (%)
	1	2	3	4	5				
CHC-UFPR	8	19	5	52	16	3.09×10^{-2}	11	21	66.9
CH-UFC	6	7	6	57	25	4.61×10^{-2}	3	5	82.2
CHU-UFPA	4	17	14	52	13	2.57×10^{-2}	19	24	64.3
HC-UFG	7	16	10	52	14	2.79×10^{-2}	15	22	66.3
HC-UFMG	4	11	5	61	19	3.66×10^{-2}	7	8	79.9
HC-UFPE	5	19	9	58	9	2.09×10^{-2}	27	20	67.3
HC-UFTM	8	11	11	56	14	2.85×10^{-2}	14	18	70.1
HC-UFU	8	17	11	53	11	2.34×10^{-2}	21	25	63.6

Table 5. *Cont.*

Hospital	Likert Scale (% of Evaluations)					PMax	Rank PMax	Rank Ebserh	Ebserh Result (%)
	1	2	3	4	5				
HDT-UFT	0	12	6	82	0	1.21×10^{-2}	35	10	77.8
HE-UFPEL	22	31	3	25	19	3.37×10^{-2}	8	34	44.4
HUAB-UFRN	6	8	12	54	21	3.86×10^{-2}	5	15	72.2
HUAC-UFCG	0	12	5	67	17	3.33×10^{-2}	10	6	81.4
HUAP-UFF	10	25	7	50	9	1.99×10^{-2}	30	29	57.9
HUB-UnB	4	21	8	53	14	2.73×10^{-2}	16	23	65.3
HUCAM-UFES	3	9	6	59	23	4.27×10^{-2}	4	7	81.4
HU-FURG	13	19	9	44	16	3.00×10^{-2}	12	28	59.4
HUGD-UFGD	17	28	10	38	7	1.59×10^{-2}	33	35	44.1
HUGG-Unirio	8	23	15	48	8	1.78×10^{-2}	31	31	55
HUGV-UFAM	11	24	9	51	5	1.47×10^{-2}	34	32	54.3
HUJB-UFCG	0	11	0	78	11	2.65×10^{-2}	17	1	88.9
HUJM-UFMT	7	11	4	69	9	2.30×10^{-2}	22	9	78
HUL-UFS	19	38	10	29	5	1.15×10^{-2}	36	36	31.8
HULW-UFPB	5	17	5	58	15	2.99×10^{-2}	13	12	72.7
HUMAP-UFMS	7	17	3	64	9	2.22×10^{-2}	23	13	72.5
HUOL-UFRN	6	16	10	55	13	2.65×10^{-2}	18	19	67.8
HUPAA-UFAL	6	15	6	65	9	2.20×10^{-2}	24	16	71.4
HUPES-UFBA	15	22	6	51	7	1.77×10^{-2}	32	30	57
HUSM-UFSM	2	9	1	62	26	4.82×10^{-2}	2	2	86.3
HU-UFJF	3	9	7	64	17	3.37×10^{-2}	9	11	77.3
HU-UFMA	14	23	8	45	10	2.13×10^{-2}	26	33	54.1
HU-UFPI	3	7	8	62	20	3.82×10^{-2}	6	4	82.5
HU-UFS	1	17	8	64	9	2.17×10^{-2}	25	14	72.4
HU-UFSC	7	22	7	56	9	2.06×10^{-2}	29	26	63
HU-UFSCar	17	0	0	50	33	6.52×10^{-2}	1	3	83.3
HU-UNIVASF	2	22	13	50	13	2.50×10^{-2}	20	27	60.4
MCO-UFBA	8	21	0	63	8	2.07×10^{-2}	28	17	70.8

6. Conclusions

The purpose of this article was to illustrate with examples the advantages of CPP in multicriteria decision aid problems that have used databases measured on Likert scales. Criticisms and controversies about the use of descriptive statistics with Likert scales have fueled the debate for decades in the scientific literature; however, CPP offers a model with full adherence to the characteristics of an ordinal scale, without involving any of the points raised for and against the use of data with that type of scale. CPP can be used with empirical distributions that are nonparametric, or for data adjustment without the need to resort to data parameters, using only the histogram. Two numerical examples and a real case were used to demonstrate the usefulness and advantages of the proposal.

The research was guided by three questions. The first aimed to highlight the main controversies in the literature surrounding the use of Likert scales for decision support. This issue was discussed in Section 2. It was demonstrated how proponents and opponents have maintained their positions for decades. This highlighted the significance of developing a strategy that does not oppose either party. The second question asked how the use of CPP could strengthen analyses based on the use of Likert scales. Methodological changes in CPP that enhance the use of Likert scales were detailed. The changes were graphically illustrated in Section 4. This contributed to the comprehension of the whole method, which was introduced in Section 3. The third question addressed the advantages of the proposed approach. Its benefits were demonstrated in three different applications presented in Section 5.

The main limitation of the proposal is the complexity of calculations that involve any mathematical model compared to the simple use of means, standard deviations, and other statistical measures from Likert scales. On the other hand, this disadvantage was mitigated by publishing the proposed software as open source on the Zenodo.org platform [29] and making it available in Appendix A. The research also did not advance on the type of management decisions that could result from the new ranking of hospitals, as it is not included in its scope.

The code in R language was used to calculate the most important parameters of the proposal, namely the probabilities of preference for the alternatives in the criteria. To complete the CPP application, it is necessary to collect data pertaining to multiple criteria and compose the probabilities according to some rule. In this article, we did not extend the analysis to this last stage, instead identifying the literature on different forms and models of CPP for this conclusion, including the "CPP" package of the R software.

A possible deepening of this study is utilizing other nonparametric distributions, such as kernel density models, and comparing the results with those of the empirical distribution explored here.

Author Contributions: Conceptualization, L.O.G. and A.P.S.; methodology, L.O.G. and A.P.S.; software, L.O.G., A.P.S., G.B.A.L. and P.A.d.A.G.; validation, L.O.G., A.P.S., G.B.A.L. and P.A.d.A.G.; formal analysis, L.O.G. and A.P.S.; investigation, L.O.G.; resources, L.O.G., A.P.S., G.B.A.L. and P.A.d.A.G.; data curation, L.O.G. and A.P.S.; writing—original draft preparation, L.O.G. and A.P.S.; writing—review and editing, L.O.G., A.P.S., G.B.A.L. and P.A.d.A.G.; visualization, L.O.G.; supervision, A.P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Probabilistic Preferences in Likert Scales with Empirical Distributions.

This R-code is intended for multicriteria decision support problems solved by the Composition of Probabilistic Preferences (CPP). The two functions calculate the joint probabilities of alternatives maximizing (PMax) and minimizing (PMin) their preferences in a criterion. The measures of the problem's decision matrix are the frequencies of responses to the Likert scale values used in the questionnaires.

This R-code is registered (DOI: 10.5281/zenodo.7950538) [29]

```
# Function 1: joint probabilities of an alternative maximize preferences by criterion
PMax.Emp.Likert = function (values,probs) {
require(mc2d)
PMax = rep(0,nrow(probs))
for (i in 1:nrow(probs)) {
PMax[i] = (integrate(Vectorize(function(x) {prod(pempiricalC(x, min(values), max(values), values, prob = probs[−i,])) * dempiricalC(x, min(values), max(values), values, prob = probs[i,])}), min(values) − 3, max(values) + 3))$value}
PMax
r = rank(−PMax)
Result = list(PMax = PMax, Rank = r)
Result}
# Function 2: joint probabilities of an alternative minimize preferences by criterion
PMin.Emp.Likert = function (values,probs) {
require(mc2d)
PMin = rep(0,nrow(probs))
```

```
for (i in 1:nrow(probs)){
PMin[i] = (integrate(Vectorize(function(x) {prod(1-pempiricalC(x, min(values), max(values), values, prob = probs[−i,])) * dempiricalC(x, min(values), max(values), values, prob = probs[i,])}), min(values) − 3, max(values) + 3))$value}
PMin
r = rank(−PMin)
Result = list(PMin = PMin, Rank = r)
Result}
# Example:
values = 1:5 # equidistant values of the Likert scale, used to evaluate alternatives
prob.a = c(2, 7, 4, 3, 4) # frequency of responses for each Likert scale value
prob.b = c(6, 3, 3, 1, 7)
prob.c = c(4, 4, 5, 2, 5)
prob.d = c(4, 4, 3, 6, 3)
prob.e = c(3, 2, 8, 6, 1)
probs = rbind(prob.a, prob.b, prob.c, prob.d, prob.e) # matrix of frequencies and values
PMax.Emp.Likert (values,probs)
PMin.Emp.Likert (values,probs)
```

References

1. Sant'Anna, A.P. *Probabilistic Composition of Preferences, Theory and Applications*, 1st ed.; Springer: New York, NY, USA, 2015.
2. Gavião, L.O.; Sant'Anna, A.P.; Garcia, P.A.A.; Silva, L.C.E. Multi-criteria decision support to criminology by Graph Theory and Composition of Probabilistic Preferences. *Pesqui. Oper.* **2021**, *41*, e249751. [CrossRef]
3. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.* **1932**, *22*, 5–55.
4. Li, Q. A novel Likert scale based on fuzzy sets theory. *Expert Syst. Appl.* **2012**, *40*, 1609–1618. [CrossRef]
5. Anjaria, K. Knowledge derivation from Likert scale using Z-numbers. *Inf. Sci.* **2022**, *590*, 234–252. [CrossRef]
6. Karuppiah, K.; Sankaranarayanan, B. An integrated multi-criteria decision-making approach for evaluating e-waste mitigation strategies. *Appl. Soft Comput.* **2023**, *144*, 110420. [CrossRef]
7. Alshamsi, A.M.; El-Kassabi, H.; Serhani, M.A.; Bouhaddioui, C. A multi-criteria decision-making (MCDM) approach for data-driven distance learning recommendations. *Educ. Inf. Technol.* **2023**, *28*, 1–38. [CrossRef]
8. Daniel, M.; Ahammed, M.M.; Shaikh, I.N. Selection of Greywater Reuse Options Using Multi-criteria Decision-making Techniques. *Water Conserv. Sci. Eng.* **2023**, *8*, 2. [CrossRef]
9. Wang, S.; Li, L.; Liu, C.; Huang, L.; Chuang, Y.-C.; Jin, Y. Applying a multi-criteria decision-making approach to identify key satisfaction gaps in hospital nurses' work environment. *Heliyon* **2023**, *9*, e14721. [CrossRef]
10. Stevens, S.S. On the Theory of Scales of Measurement. *Science* **1946**, *103*, 677–680. [CrossRef]
11. Williams, M.N. Levels of measurement and statistical analyses. *Meta-Psychology* **2021**, *5*, 1–14. [CrossRef]
12. Bishop, P.A.; Herron, R.L. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int. J. Exerc. Sci.* **2015**, *8*, 297–302.
13. Tanujaya, B.; Prahmana, R.C.I.; Mumu, J. Likert scale in social sciences research: Problems and difficulties. *FWU J. Soc. Sci.* **2022**, *16*, 89–101.
14. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 8th ed.; Cengage Learning EMEA: London, UK, 2018.
15. Kuzon, W.; Urbanchek, M.; McCabe, S. The seven deadly sins of statistical analysis. *Ann. Plast. Surg.* **1996**, *37*, 265–272. [CrossRef]
16. Wu, H.; Leung, S.-O. Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *J. Soc. Serv. Res.* **2017**, *43*, 527–532. [CrossRef]
17. Awang, Z.; Afthanorhan, A.; Mamat, M. The Likert scale analysis using parametric based Structural Equation Modeling (SEM). *Comput. Methods Soc. Sci.* **2016**, *4*, 13–21.
18. Baran, T. Comparison of parametric and non-parametric methods to analyse the data gathered by a likert-type scale. In *Handbook of Research on Applied Data Science and Artificial Intelligence in Business and Industry*; IGI Global: Hershey, PA, USA, 2021; pp. 414–430.
19. Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* **2010**, *15*, 625–632. [CrossRef]
20. Harpe, S.E. How to analyze Likert and other rating scale data. *Curr. Pharm. Teach Learn* **2015**, *7*, 836–850. [CrossRef]
21. Thomas, M.A. Mathematization, Not Measurement: A Critique of Stevens' Scales of Measurement. *J. Methods Meas. Soc. Sci.* **2019**, *10*, 76–94. [CrossRef]
22. Mircioiu, C.; Atkinson, J. A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy* **2017**, *5*, 26. [CrossRef]
23. Pornel, J.B.; Saldaña, G.A. Four common misuses of the Likert scale. *Philipp. J. Soc. Sci. Humanit.* **2013**, *18*, 12–19.
24. Jamieson, S. Likert scales: How to (ab) use them? *Med. Educ.* **2004**, *38*, 1217–1218. [CrossRef] [PubMed]

25. Allen, I.E.; Seaman, C.A. Likert scales and data analyses. *Qual. Prog.* **2007**, *40*, 64–65.
26. Sullivan, G.M.; Artino, A.R., Jr. Analyzing and interpreting data from Likert-type scales. *J. Grad. Med. Educ.* **2013**, *5*, 541–542. [CrossRef] [PubMed]
27. Alabi, A.T.; Jelili, M.O. Clarifying likert scale misconceptions for improved application in urban studies. *Qual. Quant.* **2023**, *57*, 1337–1350. [CrossRef]
28. Schrum, M.L.; Ghuy, M.; Hedlund-Botti, E.; Natarajan, M.; Johnson, M.J.; Gombolay, M.C. Concerning Trends in Likert Scale Usage in Human-robot Interaction: Towards Improving Best Practices. *ACM Trans. Human-Robot Interact.* **2023**, *12*, 1–32. [CrossRef]
29. Gavião, L.O.; Sant′Anna, A.P.; Lima, G.B.A.; Garcia, P.A.A. *Probabilistic Preferences of Likert Scale Data by Empirical Distributions*; Version 2.0.; Zenodo.org: Geneve, Switzerland, 2023; p. 1.
30. Sant′Anna, A.P.; Sant′Anna, L.A.F.P. Randomization as a stage in criteria combining. In Proceedings of the International Conference on Industrial Engineering and Operations Management—VII ICIEOM, Salvador, Brazil, 4–8 October 2001; pp. 248–256.
31. Gavião, L.O.; Sant'Anna, A.P.; Lima, G.B.A.; de Almada Garcia, P.A.; de Sousa, A.M. Selecting distribution centers in disaster management by Network Analysis and Composition of Probabilistic Preferences. In *Industrial Engineering and Operations Management*; Amorim, A.M.T.T., Barbastefano, R.G., Scavarda, L.F., Reis, J.C.G.D., Amorim, M.P.C., Eds.; PUC-RJ: Rio de Janeiro, Brazil, 2020; pp. 1–11.
32. de Souza, F.H.; Gavião, L.O.; Sant'Anna, A.P.; Lima, G.B. Prioritizing risks with composition of probabilistic preferences and weighting of FMEA criteria for fast decision-making in complex scenarios. *Int. J. Manag. Proj. Bus.* **2021**, *15*, 572–594. [CrossRef]
33. Sant′anna, A.P.; Gavião, L.O.; Lima, G.B.A. Alternatives for the composition of interactive environmental impact factors. *Pesqui. Oper.* **2022**, *42*, e247786. [CrossRef]
34. Sant′anna, A.P.; Gavião, L.O.; Sant'anna, T.L. Multi-criteria classification of reward collaboration proposals. *IISE Trans.* **2023**. [CrossRef]
35. Gavião, L.O.; Sant′anna, A.P.; Lima, G.B.A.; Garcia, P.A.A. Evaluation of soccer players under the Moneyball concept. *J. Sports Sci.* **2019**, *38*, 1221–1247. [CrossRef]
36. Gavião, L.O.; Meza, L.A.; Lima, G.B.; Sant'Anna, A.P.; de Mello, J.C.B.S. Improving discrimination in efficiency analysis of bioethanol processes. *J. Clean. Prod.* **2017**, *168*, 1525–1532. [CrossRef]
37. Gavião, L.O.; Sant′Anna, A.P.; Lima, G.B.A.; Garcia., P.A.A. *CPP: Composition of Probabilistic Preferences*; R Package Version 0.1.0.; R Foundation for Statistical Computing: Vienna, Austria, 2022; pp. 1–24.
38. Wandresen, R.R.; Netto, S.P.; Koehler, H.S.; Sanquetta, C.R.; Behling, A. Nonparametric method: Kernel density estimation applied to forestry data. *Floresta* **2019**, *49*, 561–570. [CrossRef]
39. Jiang, T.; Li, D. Approximation of Rectangular Beta-Laguerre Ensembles and Large Deviations. *J. Theor. Probab.* **2013**, *28*, 804–847. [CrossRef]
40. Chalabi, Y.; Scott, D.J.; Würtz, D. The Generalized Lambda Distribution as an Alternative to Model Financial Returns. Inst Für Theor Phys Univ Auckland: Zürich, Auckland. Available online: www.Rmetrics.Org/Sites/Default/Files/Glambda.Pdf; www.rmetrics.org/sites/default/files/2009-01-glambdaDist.pdf (accessed on 30 June 2023).
41. Zhou, Z.; Azam, S.S.; Brinton, C.; Inouye, D.I. Efficient Federated Domain Translation. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023; pp. 1–31.
42. Pouillot, R.; Delignette-Muller, M.L. Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *Int. J. Food Microbiol.* **2010**, *142*, 330–340. [CrossRef]
43. Garcia, P.A.A.; Sant'Anna, A.P. Vendor and logistics provider selection in the construction sector: A Probabilistic Preferences Composition approach. *Pesqui. Oper.* **2015**, *35*, 363–375. [CrossRef]
44. Sant′anna, A.P.; Meza, L.A.; Ribeiro, R.O.A. Probabilistic composition in quality management in the retail trade sector. *Int. J. Qual. Reliab. Manag.* **2014**, *31*, 718–736. [CrossRef]
45. Gaviao, L.O.; Sant′Anna, A.P.; Lima, G.B.A.; Garcia, P.A.A.; Kostin, S.; Asrilhant, B. Selecting a Cargo Aircraft for Humanitarian and Disaster Relief Operations by Multicriteria Decision Aid Methods. *IEEE Trans. Eng. Manag.* **2019**, *67*, 631–640. [CrossRef]
46. Sant′Anna, A.P.; Faria, F.; Costa, H.G. Aplicação da Composição Probabilística e do método das K-Médias à classificação de municípios quanto à oferta de creches. *Cad. Do IME-Série Estatística* **2013**, *34*, 17.
47. Gavião, L.O.; Silva, R.F.d.; Sant'Anna, A.P.; Lima, G.B.A. *Ordenação de Municípios por Potencial de Contaminação de Águas com Fármacos Oncológicos por Composição Probabilística de Preferências.*; XLVIII Simpósio Brasileiro de Pesquisa Operacional: Vitória, Brazil, 2016; p. 12.
48. Garcia, P.A.A.; Garcia, V.S.; Saldanha, P.L.C.; Jacinto, C.M.C. Combined use of composition of probabilistic preferences and entropy weighting for failure mode prioritization. In Proceedings of the European Safety and Reliability Conference—ESREL, Zurich, Switzerland, 7–10 September 2015; CRC Press: Boca Raton, FL, USA, 2015; pp. 381–386.