Article title: Comparing Discrete and Continuous Data: Concepts, Differences, and Applications.
Authors: Akpusugh Verse Ephraim Odvese[1], Dekpoghol Tertsea[2]
Affiliations: department of mathematics, federal university of technology minna, nigeria.[1], department of chemistry, university of abuja, fct nigeria.[2]
Orcid ids: 0009-0002-9520-4868[1], 0009-0008-6399-2314[2]
Contact e-mail: akpusughephraim@gmail.com

DOI: 10.14293/PR2199.001249.v1
Preprint first posted online: 01 November 2024
Keywords: Discrete data, Continuous data, Bar charts, Pie charts, Data Visualization, Mean, Median , Standard Deviation , Data analysis, Whole numbers

Comparing Discrete and Continuous Data: Concepts, Differences, and Applications.

Akpusugh V. E.[1] & Dekpoghol, T.[2]

[1]Department of Mathematics, Faculty of Physical Sciences, Federal University of Technology Minna.

[2]Department of Chemistry, University of Abuja, Federal Capital Territory.

Corresponding mail: akpusughephraim@gmail.com

Abstract

Data is usually divided into two types: discrete and continuous, both of which play important roles in subjects such as statistics, mathematics, computer science, and data analysis. Understanding the distinctions between discrete and continuous data is critical for selecting proper data analysis techniques, statistical tests, and graphical representations. Discrete data shows countable values or separate categories that are often derived from a finite set, such as integers or category labels. This data type is distinguished by solitary points and is frequently found in survey replies, customer counts, or categorical variables (for example, colors types). Discrete data is commonly represented using bar charts or pie charts, and analysis are frequently performed using chi-square tests or other non-parametric approaches that account for discrete distributions. Continuous data, on the other hand, includes an infinite number of possible values within a specific range, and is frequently used to represent measurable quantities such as time, temperature, height, or distance. Continuous data points are on a continuum and can take any value within a certain interval, tailoring them for expressing phenomena that fluctuate smoothly across time or space. Continuous data, due to its mathematical qualities, can be evaluated with a variety of parametric tests, including t-tests and ANOVA, and is frequently represented using histograms or line charts to show data density or trends over time. Data processing, analysis techniques, and computational complexity are all radically altered by the distinction between discrete and continuous data. Because of its finite nature, discrete data frequently calls for simpler computing models and storage, whereas continuous data usually necessitates more intricate data handling and modeling procedures because of its infinite granularity. This paper explores the differences between discrete and continuous data in data science and machine learning, giving prominence to selecting appropriate algorithms for each type. It emphasizes the advantages and disadvantages of each type in various applications and provides recommendations for data handling and analysis techniques.

Keyword: Discrete data, Continuous data, Bar charts, Pie charts, Data visualization, Mean, Median, Standard deviation, Data analysis, Whole numbers, Line graphs, Decimals.

Introduction

Identifying and interpreting data types is a critical stage in data analysis and statistical studies, as it defines the entire analytical process. Data kinds influence how information is acquired, stored, and processed, influencing the accuracy and usefulness of the key points of data. Each type of data requires its own handling and analysis techniques, therefore researchers, analysts, and decision-makers must be able to discern between them accurately. Discrete and continuous data are the two types of data that play a major role in structural analysis, with each having its own set of qualities, applications, and implications (Khetwal, 2023).

Different, countable variables, such as the number of students in a classroom or the quantity of commodities sold, comprise discrete data. Counts, proportions, and classification methods are typically the most effective ways to analyze this data type, which depicts quantities that can only take specific values with gaps in between. Continuous data, which depicts quantifiable quantities such as temperature, height, or time, has an indefinite range of values inside a defined interval. Because values can be indefinitely separated, this type of data provides for much higher resolution. It is widely researched using statistical approaches that require accurate measurement, such as determining averages, ranges, and fitting continuous curves (Delignette, 2015). Because it directly influences decisions about data collecting strategies, statistical approaches, and result interpretation, the ability to distinguish between discrete and continuous data is essential for efficient data analysis. Bar charts, for instance, are best suited for discrete data, but line graphs or histograms are more appropriate for continuous data (Alstott, 2014). Misunderstanding the nature of data can result in incorrect procedures, biased analysis, and deceptive findings, all of which can have practical repercussions, particularly in industries like scientific research, healthcare, and finance (Shen, 2014).

This article's goal is to give a thorough comparison between discrete and continuous data, talking about their key distinctions, defining traits, and real-world uses across a range of industries. More precise and perceptive data-driven decision-making across domains will eventually be possible thanks to this comparative understanding, which will act as a guide for choosing the best analytical methods for each type of data (Leibovich, 2014). Since it directly affects how data is gathered, examined, and interpreted, knowing the difference between discrete and continuous data is vital to data analysis. Certain techniques, such counts, proportions, or frequency distributions, are needed for the collection and analysis of discrete data, which is made up of discrete, countable values (Capó, 2020). These techniques are frequently appropriate for data that belongs to categories or entire numbers. Contrariwise, continuous data encompasses an infinite range of values within predetermined intervals, enabling more precise measurement and analysis methods like curve-fitting, variability, and central tendency measurements (Schöllhorn, 2002).

Making this distinction is essential because using the incorrect techniques on a particular sort of data might produce distorted results, misunderstandings, or even incorrect conclusions. In order to ensure exact and unambiguous precisions, continuous data is better represented by line graphs or histograms, but discrete data is typically best displayed by bar charts (De Leon, 2011). Inaccurate identification and application of the proper methodologies for each type may cause analysis to miss the real patterns in the data. With a thorough comparison of discrete and continuous data, this article explores their distinct qualities, distinctions, and uses in a variety of sectors, giving readers a better knowledge to improve the precision and interpretive clarity of their data analysis (Liu, 2019). Through an examination of their unique features, distinctions, and real-world applications, this article aims to give readers a comparative grasp of discrete versus continuous data (Kaur, 2021). The article seeks to shed light on the definitions and applications of these two fundamental data kinds, providing guidance on how to handle, analyze, and understand data according to its type. By emphasizing the advantages and disadvantages of discrete and continuous data, this comparison aims to help readers choose appropriate analytical methods and strategies for different situations, thereby improving the precision and comprehensiveness of data-driven choices (Cartuyvels, 2021).

Statement of the Problem

This article addresses a typical difficulty in data analysis and statistical studies: understanding and using discrete and continuous data effectively. Many specialists, researchers, and analysts struggle to distinguish between these two types of data, which can lead to wrong interpretations, flawed analysis, and poor decision-making. Ignoring the distinctions between discrete and continuous data collection, analysis, and presentation might result in erroneous conclusions or wasteful solutions. This article seeks to address this issue by providing readers with a clear, comparative understanding of discrete and continuous data, its unique qualities, and appropriate applications.

Objectives of the Study

The purpose of this research is to elucidate the basic traits of discrete and continuous data, showing how each kind is defined and outlining its distinct qualities.

i. This study compares discrete and continuous data to show how each type influences data collection, statistical method selection, and result interpretation.
ii. Analytical approaches that improve data analysis accuracy and understanding will be emphasized as the study looks at appropriate approaches for both discrete and continuous data.
iii. This study intends to offer recommendations for the appropriate representations, such as line graphs or histograms for continuous data and bar charts for discrete data, in order to guarantee precise and efficient data display and interpretation.

Methodology

Descriptive statistics and visualization are employed in this study's technique to emphasize the variations in traits and behavior between the two categories.

Table 1: Key differences between discrete data and continuous date

Feature	Discrete Data	Continuous Data
Definition	Data that can take on a finite number of values	Data that can take on an infinite number of values within a range
Examples	Number of students, cars, or pets	Height, weight, temperature
Measurement	Usually counted	Usually measured
Data Type	Integer or whole numbers	Real numbers (including fractions)
Graphical Representation	Bar charts or pie charts	Line graphs or histograms
Gaps	Contains gaps between values	No gaps between values
Mathematical Operations	Can only perform certain operations (i.e., addition)	Can perform a wide range of operations (addition, subtraction, multiplication, etc.)

Table 2: Discrete Data: Daily Customer Count at a Restaurant

Day	Number of Customers
Monday	45
Tuesday	52
Wednesday	48
Thursday	50
Friday	60
Saturday	78
Sunday	65

Table 3: Continuous Data: Daily Temperature Readings at 12 PM (°C)

Day	Temperature (°C)
Monday	23.4
Tuesday	22.1
Wednesday	23.9
Thursday	24.5
Friday	25.2
Saturday	27.3
Sunday	26.1

Data Distribution

Data (Customer Count): The data values are entire, distinct numbers that correspond to daily consumer counts. There are no intermediate values in the distribution; instead, there are obvious jumps between numbers (such as 45, 52, 4).

Continuous Data (Temperature): Decimal numbers are permitted since data values fall within a continuous range (e.g., 23.4, 22.1). Trends are easy to spot because of the distribution's smoothness and daily variations.

Range and Granularity

Discrete Data: The number of customers varies between 45 and 78. There are no values in between because this sort of data only records full numbers (for example, 51.5 customers cannot exist).

Continuous Data: 22.1°C to 27.3°C is the temperature range, with decimal accuracy. This granularity makes it possible to make subtle differences and detect even minute temperature changes.

Central Tendency Measures

Mean (Average)

Customers:

$$\frac{45+52+48+50+60+78+65}{7}$$

= 56.14 customers

Temperature:

$$\frac{23.4+22.1+23.9+24.5+25.2+27.3+26.1}{7}$$

= 24.64°C

Median

Customers: 52 is the median because it is the midpoint value in an ordered set of numbers.

Temperature: In terms of ordered temperatures, the median is 24.5°C.

Dispersion and Variability

Range:

Customers:

78 - 45 = 33 customers

Temperature:

27.3 - 22.1 = 5.2°C

Standard Deviation:

The standard deviation of discrete data, such as the number of customers, will be higher because the numbers differ more.

The smoother range of values in continuous data (temperature) will be reflected in a smaller standard deviation.

The formula adopted in determining the sample standard deviation (*S*) is given as;

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Where;

N = Number of data points

X_i = Each individual data point

$\bar{X}$ = Mean of the data set

S = Sample standard deviation

Applying the formula to both the customer count and the temperature sets of data, the standard deviation for each set of data is as follows;

Customer Count: The daily customer count varies greatly depending on the day of the week, as seen by the larger standard deviation of 11.64.

Temperature: With a 1.74 standard deviation, there is little daily variation, suggesting that temperatures were mostly constant.

Visual Comparison

Several chart formats are very useful for discrete data, which consists of discrete, distinct numbers (often counts or categories):

i. Bar Chart: shows distinct categories with rectangular bars, each bar's length representing the category's value. This is among the most used methods for displaying discrete data.

ii. Pie Chart: shows, as pie slices, the percentage of each category. The amount of each slice represents the category's contribution to the total, despite the fact that pie charts are frequently criticized for being less useful for comparison than bar charts.

iii. Column Chart: With a vertical orientation, it resembles a bar chart. A category is represented by each column, and the height denotes the value.

iv. Dot Plot: allows for a rapid visual comparison of discrete values by displaying individual data points along a number line.

v. Frequency Table: This tabular representation, which is useful for analysis, displays the number of occurrences for each category even though it is not a graphical chart.

vi. Stacked Bar Chart: Comparable to a bar chart, but with segments that correspond to subcategories, each bar indicates a total, enabling comparison of individual components to the whole.

Bar Chart for Discrete Data (Customer Count): This will show individual bars representing each day's customer count, with clear differences between days.

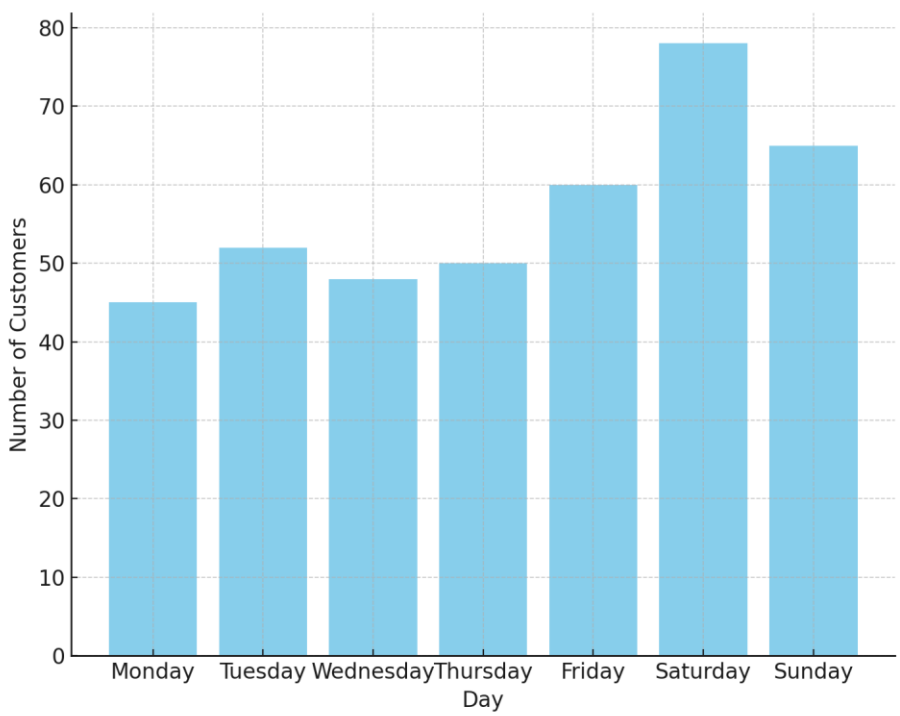

Figure 1: Daily customer count at a restaurant

The diagram above is a bar chart that displays the number of persons at the restaurant each day, emphasizing variations in traffic throughout the week. The customer count bar chart highlights a clear increase in customers as the week progresses, with a peak on the weekend. This trend is useful for identifying high-demand days.

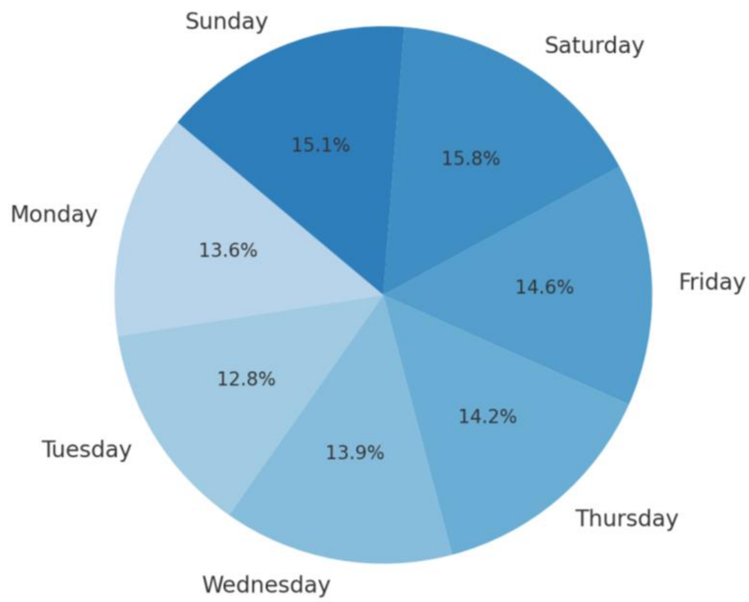

Figure 2: Proportion of daily temperatures (ºC)

The diagram above is a pie chart that shows the percentage of each day's temperature in relation to the weekly total, providing summary of the distribution of temperatures. The temperature pie chart provides perception into each day's contribution to the week's overall temperature distribution. Although there is little difference in daily temperatures, Saturday contributes the highest portion.

When dealing with continuous data, charts that can depict a range of values along a continuum are usually used. The following are typical chart types for continuous data:

i. Histogram: illustrates how data is distributed over intervals (bins) by displaying the frequency distribution of continuous data.

ii. Line Chart**:** is helpful for showing how data evolves or changes over time by presenting trends or continuous data points.

iii. Scatter Plot**:** Plotting data points on an x-y axis is a common method for observing correlations or relationships between two continuous variables.

iv. Box Plot (Box-and-Whisker Plot)**:** uses quartiles to summarize the data distribution, emphasizing the median, dispersion, and any outliers.

v. Density Plot: A histogram that has been smoothed out to display the probability distribution of continuous data.

vi. Area Chart**:** Useful for displaying cumulative data or trends over time, it resembles a line chart but fills the space beneath the line.

Line Chart for Continuous Data (Temperature): A line graph for temperature will show a smooth curve, making it easier to identify a trend (e.g., temperature increasing toward the weekend).

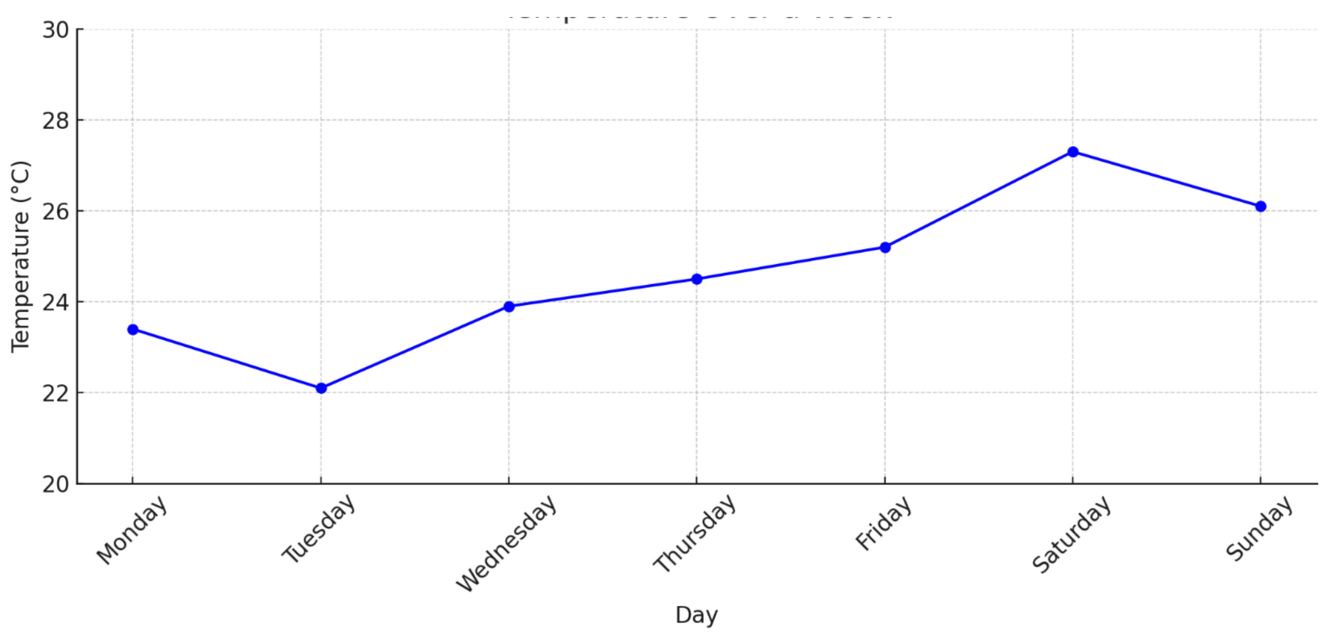

Figure 3: Temperature over a week

The figure above is a line chart representing the temperature over the week. The chart shows the temperatures for each day, illustrating how they change from Monday to Sunday.

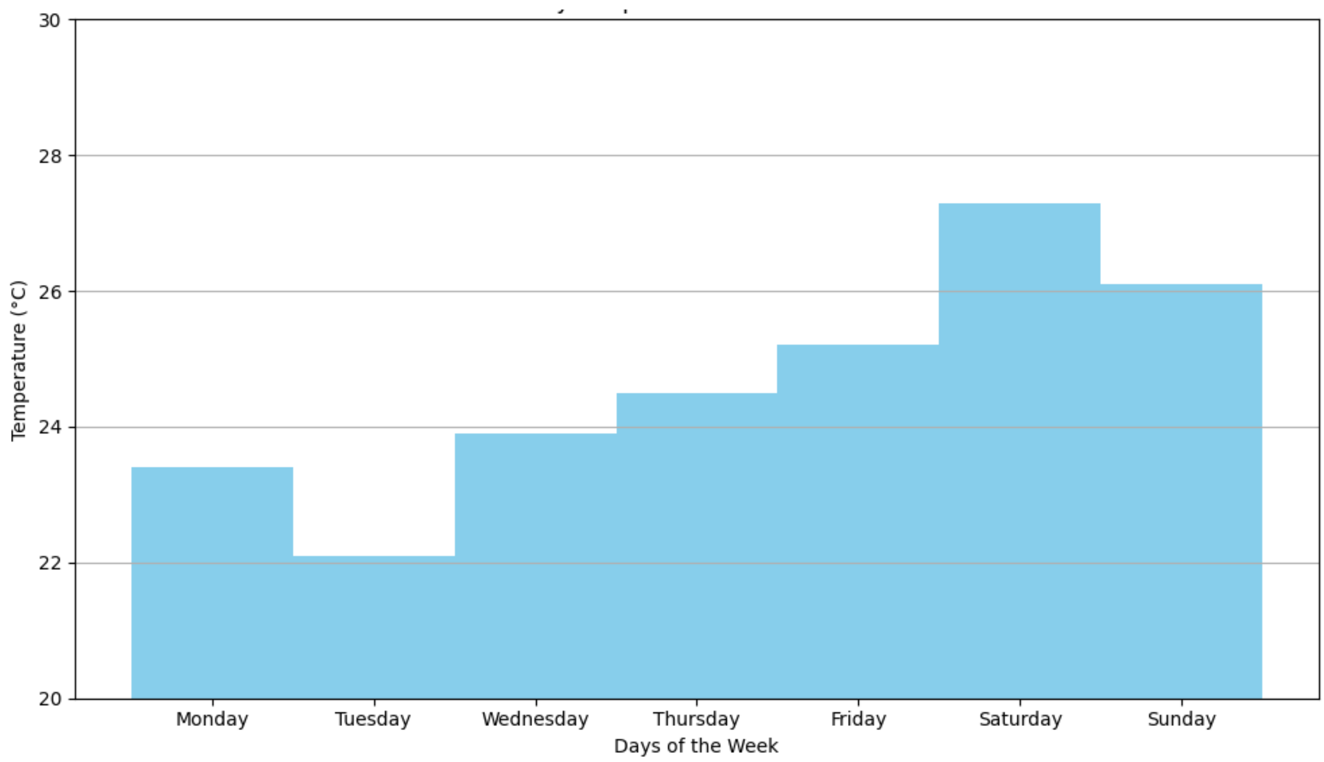

Figure 4: Daily temperatures for the week (ºC)

The figure above is a histogram showing the daily temperatures for the week. Each bar represents the temperature for a specific day, with the values ranging from around 22°C to 27°C.

Results and Discussion

The summary of the data simulated in the study is given below;

Discrete Data (Customer Count): With a weekly average of roughly 56.14 customers, the customer count exhibits notable daily fluctuations. The number of clients was at its highest on Saturday (78) and at its lowest on Monday (45).

Continuous Data (Temperature**)**: With only a 5.2°C range of variation, the mean daily temperature for the week was a comparatively constant 24.64°C. Tuesday was the coolest day (22.1°C), and Saturday was the warmest (27.3°C).

The following discussions can be made from the study

i. Discrete Data (Customer Count): This data type is ideal for analyzing count-based, categorical trends and patterns. Here, the variability in customer count could suggest factors like work schedules or weekend preferences influencing visits to the coffee shop. High standard deviation and distinct peaks highlight these day-based patterns.

ii. Continuous Data (Temperature): Continuous data suits contexts where precise measurements matter, as shown in temperature variations. Although day-to-day temperatures didn't vary noticeably, continuous tracking helps detect gradual trends, useful in seasonal forecasting or climate studies.

The following practical implications were been derived from the study

i. Staffing and Inventory for the restaurant: Managers can prepare to boost staffing and supplies in accordance with the knowledge that customer traffic increases on weekends. Reduced foot traffic early in the week indicates a chance to increase midweek visits through loyalty programs or promotions.

ii. Temperature precisions: Although temperatures stayed constant, information about slow variations can help the store stock temperature-sensitive items or maximize energy use.

The following limitations were encountered in the course of the research and mentioned below:

i. Limited Data Points: Only one week is covered by this set of data, which restricts how far the research can be applied. A more trustworthy trend analysis might be possible with a larger dataset.

ii. Effects of External Factors: Other variables that can affect customer numbers or temperature measurements, such as weather variations or promotions, are not taken into consideration in this study.

Overall, we discover that each data type has distinct advantages and setbacks that affect the way information is gathered, examined, and understood after contrasting discrete and continuous data. Countable and categorical values make up discrete data, which is perfect for situations requiring certain numbers or groups, such the quantity of things or classifications in a category. Analyzing it is simple, but it is not as precise as continuous data, which permits measurements within a range and facilitates in-depth trend and variability research. Although continuous data is more flexible and precise, it must be measured carefully and is frequently more susceptible to outliers. Continuous data uses methods like regression and t-tests to identify trends and correlations across time, whereas discrete data usually uses frequency counts or chi-square tests. Both forms of data have useful uses in many domains, and knowing how they differ from one another aids in selecting the best strategy for certain research topics, guaranteeing more precise and notable findings.

Conclusion and Recommendations

This analysis uses real-world examples of temperature readings and daily customer counts to illustrate the differences between discrete and continuous data. Customer counts, a discrete data set that reflects day-based foot traffic patterns, exhibit notable weekly variability. Continuous data, represented by temperature, displays smaller fluctuations and allows for precise trend analysis. Both data types offer unique precisions, discrete data is valuable for analyzing specific counts and demand patterns, while continuous data is essential for monitoring gradual changes and detailed measurements. The research shows how choosing the right analysis techniques, creating effective visualizations, and drawing useful conclusions all depend on knowing the nature of the data whether it be continuous or discrete. Particularly for companies with fluctuating client traffic, the results may help with staffing, inventory planning, and resource allocation.

The following recommendations are been made in order for analysts to adhere to in order to get the best results:

i. It is essential to know if your data should be continuous or discrete. Discrete data is suitable for count-based scenarios (such as the frequency of occurrences or the number of flaws). Continuous data is more suited for situations requiring accuracy and measurements (such temperature, time, or distance).

ii. Continuous and discrete data demand various statistical methods because of their varied characteristics. Poisson distributions and chi-square tests are examples of count-based methods that perform well with discrete data. Because values in continuous data are continuous, techniques like regression analysis, t-tests, and ANOVA are appropriate.

iii. For effective information presentation, choose the right charts for each type of data. Pie charts and bar charts are effective for showing counts or categories when dealing with discrete data. Scatter plots, line graphs, and histograms are better for continuous data because they make it easier to see correlations, trends, and distributions.

References

Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *PLoS ONE*, *9*(1), e85777. https://doi.org/10.1371/journal.pone.0085777

Capó, M., Pérez, A., & Lozano, J. A. (2020). An efficient K-means clustering algorithm for tall data. *Data Mining and Knowledge Discovery*, *34*(3), 776–811. https://doi.org/10.1007/s10618-020-00678-9

Cartuyvels, R., Spinks, G., & Moens, M. F. (2021). Discrete and continuous representations and processing in deep learning: Looking forward. *AI Open*, *2*, 143-159. https://doi.org/10.1016/j.aiopen.2021.07.002

De Leon, A. R., Soo, A., & Williamson, T. (2011). Classification with discrete and continuous variables via general mixed-data models. *Journal of Applied Statistics*, *38*(5), 1021-1032. https://doi.org/10.1016/j.artmed.2018.10.002

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: AnRPackage for Fitting Distributions. *Journal of Statistical Software*, *64*(4). https://doi.org/10.18637/jss.v064.i04

Kaur, H. (2021). Analysis of Various Optimization Techniques in Machine Learning. *Türk Bilgisayar Ve Matematik Eğitimi Dergisi*, *12*(2), 472–482. https://doi.org/10.17762/turcomat.v12i2.855

Khetwal, A., Einstein, H. H., & Rostami, J. (2023). Combining the CSM2020 discrete event simulation model with Decision Aid in Tunneling (DAT) to develop a robust approach for the estimation of completion time for TBM tunnels. *Tunnelling and Underground Space Technology*, *138*, 105156. https://doi.org/10.1016/j.tust.2023.105156

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *Quarterly Journal of Experimental Psychology*, 67(5), 899-917. https://doi.org/10.1080/17470218.2013.837940

Liu, M., Stella, F., Hommersom, A., Lucas, P. J., Boer, L., & Bischoff, E. (2019). A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity. *Artificial intelligence in medicine*, *95*, 104-117. https://doi.org/10.1016/j.artmed.2018.10.002

Schöllhorn, W. I., Nigg, B. M., Stefanyshyn, D. J., & Liu, W. (2002). Identification of individual walking patterns using time discrete and time continuous data sets. *Gait & Posture*, *15*(2), 180-186. https://doi.org/10.1016/S0966-6362(01)00193-X

Shen, J., & Lam, J. (2014). <inline-formula> <tex-math notation="TeX">$\ell_{\infty}/L_{\infty} $</tex-math></inline-formula>-Gain Analysis for Positive Linear Systems With Unbounded Time-Varying Delays. *IEEE Transactions on Automatic Control*, *60*(3), 857–862. https://doi.org/10.1109/tac.2014.2344295