Jurma : Jurnal Program Mahasiswa Kreatif
Universitas Ibn Khaldun Bogor, Indonesia
Vol. 8, No. 1, Juni 2024, pp. 186~191
ISSN: 2615-8019, DOI: 10.32832/pkm
http://pkm.uika-bogor.ac.id/index.php/pkm-p

A Comprehensive Guide to Understanding and Applying Descriptive Statistics in Data Analysis

Rega Priandanagari Gevisa [1], Kurniati [2]
Universitas Ibn Khaldun, Indonesia

Article Info

Article history:

Received January 31, 2024
Revised February 19, 2024
Accepted March 18, 2024

Keywords:

Descriptive
Statistics

ABSTRACT

Statistics plays a central role in parsing and interpreting data, being an important foundation, especially in scientific disciplines In this digital age, where the volume of data is constantly increasing, an understanding of statistics is essential to draw meaningful insights from the information contained in the data. Descriptive statistics, using tables and charts, provide a clear and effective way to summarize and present data. Through table types such as frequency, classification, and contingency, as well as measures of data centering and dispersion, statistics enable a deeper understanding of patterns and variations within a dataset. Therefore, mastery of statistical concepts is key to optimizing the use of data in a decision-making context.

Corresponding Author:

Rega Priandanagari Gevisa
Universitas Ibn Khaldun
Email: regagevisa@gmail.com

INTRODUCTION

In today's digital era, digital data is increasingly massive and will continue to grow. Of course, we are familiar with the terms data mining, data warehouse, machine learning, data science, and so on. Therefore, the branches il of science as mentioned above, really need statistics, you could even say that statistics is the heart of the driving force for sciences related to data.

According to Nasution, the definition of statistics develops according to the times, as follows:

1. Statistics are a set of numbers to explain something, both random numbers and numbers that have been arranged in a table.
2. Statistics is a set of methods and rules about collecting, processing, analyzing, and interpreting data consisting of numbers.
3. Statistics are a set of numbers that explain the properties of data or the results of observations/research. (Nasution, 2017)

As for descriptive statistics according to Adhiwibowo in a book entitled Introduction to Statistics: For Students and the Public, states that descriptive statistics are statistics that describe the activities of collecting, compiling, processing, and presenting data in the form of numbers through tables or diagrams, with the aim of providing a clear and concise description of the situation or event. (Adhiwibowo & Putri, 2023).

ISSN: 2615-8019

Data Presentation

Presentation of data is an activity in making research results in order to see the results of research and analyze or understand with a specific purpose. In addition, data presentation is also very useful in order to provide an overview to readers about the contents of the data presented. The data presented can be in various forms of diagrams. Existing diagrams include bar charts, line charts, and pie charts. (Adhiwibowo & Putri, 2023). In his book, he also said that diagrams can also be presented in the form of tables.

1. Table

A table is data that is presented through a combination of rows and columns. According to Nasution, based on the organization of the data, tables are divided or distinguished by several types, namely:

a. Frequency tables are used in statistics to group data into ranges or categories and show frequencies. Here is a case study with an example of a frequency table:

At SMK Kosgoro in Bogor City, selection is being held for the Paskibra competition. One of the requirements is that candidates have a minimum height of 170cm. Therefore, measurements were taken of each student in a class and the following data was obtained:

Height (cm)	Frequency
140-149	7
150-159	12
160-169	17
170-179	2
180-189	1

Based on the table above, we can see that those who meet the requirements are 3 students whose height is above 170cm.

b. A classification table is a form of table used to group and organize data based on certain criteria. The types of classification tables can vary depending on the context or domain in which they are used. Here are some examples of common classification tables:

Type	Example	Food	Habitat
Mammals	Cattle, Sheep	Herbivores	Land
Reptiles	Snake	Omnivores	Land
Birds	Eagle	Carnivores	Air

c. Contingency table, A contingency table, also known as a cross table or crosstab, is a form of table used in statistics to present the concurrent distribution of two or more variables. This table helps to see the relationship between these variables. The cross tabulation method can answer the relationship between two or more research variables but not the causal relationship. The more the number of variables tabulated, the more complex the interpretation. (Nisva & Ratnasari, 2020).

	Men	Women
Ball	20	15
Swimming	25	30

The most common example is a 2x2 contingency table, but there can be tables with more categories. Below is an example of a simple 2x2 contingency table:

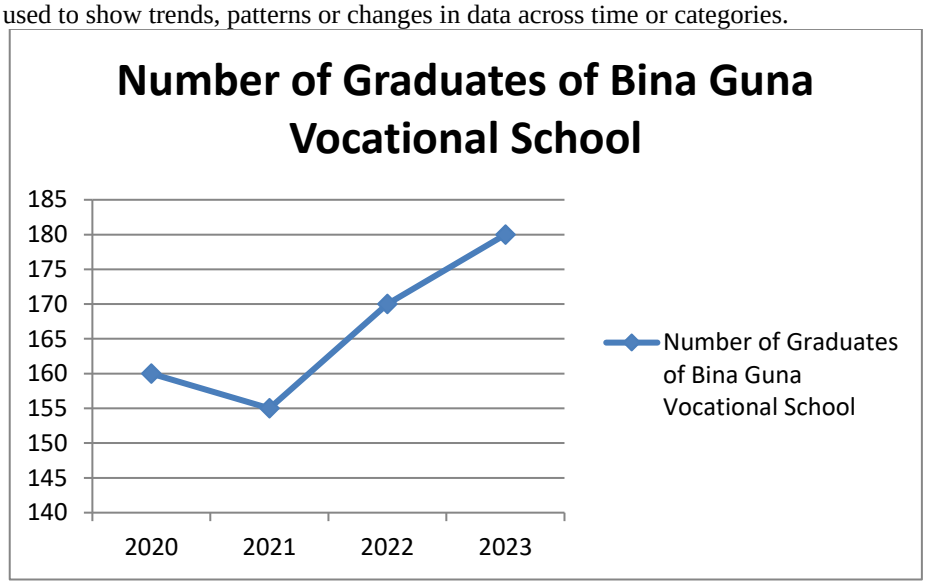
In this example, the table shows the relationship between two variables: gender (Male or Female) and hobby (Ball or Swimming). The numbers in the cells represent the number of individuals who belong to a particular combination of those categories.

The use of contingency tables can help in statistical analysis, especially in the context of chi-square tests to determine if there is a significant relationship between two variables. The table is also useful for providing a clear picture of the distribution and relationship of the variables.

d. Correlation table, which is a table that shows or contains the correlation (relationship) between the data presented.

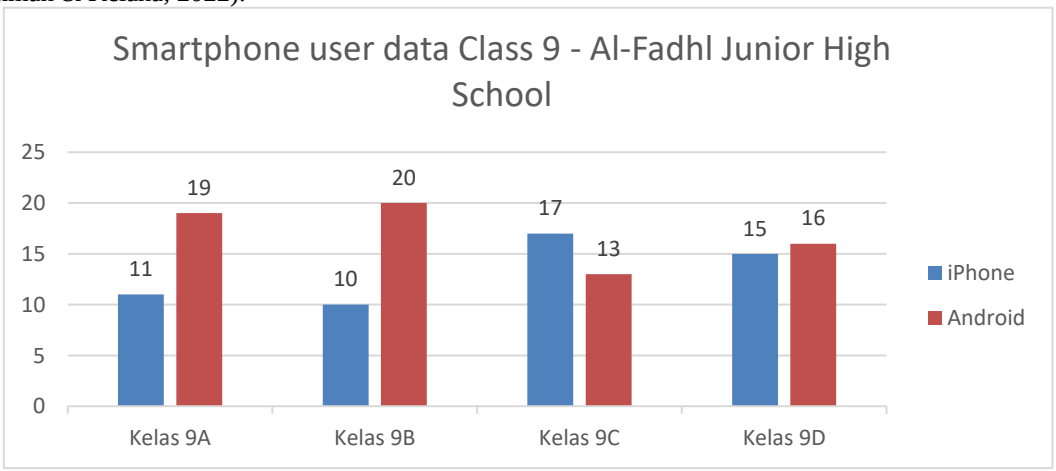
2. Line Diagram

A type of graph that uses lines to connect data points in a specific time series or sequence. They are commonly used to show trends, patterns or changes in data across time or categories.



3. Bar Diagram

Bar chart is a form of diagram that represents data in the form of vertical or horizontal rectangles. (Rochman & Kelana, 2022).



4. Circle Diagram

There are two types of pie charts, one is a pie chart that uses degrees, then the other is a diagram that uses percentages. using percentages. However, generally and often encountered is a diagram with percentages. Here is an example of a pie chart with a case study:

As of August 2023, Vision Publisher has sold 4 titles books with the following sales details: Book A as many as 325 copies, book B as many as 185 copies, book C as many as 278 copies, and book D as many as 90 copies. copies, and book D as many as 90 copies. Here is the diagram circle diagram obtained:

Data Centering Measures

A measure of data centering is a statistic used to measure the "center" or middle value of a data set. This measure provides an overview of the location of values in the data. Some commonly used measures of data centering include:

1. Mean

The average is the sum of all values in a data set divided by the number of values. The average is often used as a common measure of center because it is sensitive to all values in the data.

$$\text{Mean} = \frac{Sum\ of\ all\ values}{Total\ value}$$

ISSN: 2615-8019

2. Median

The median is the middle value in a sorted data set. If the number of data is odd, the median is the middle value. If the number of data is even, the median is the average of the two center values.

3. Mode

A mode is the value or values that appear most frequently in a data set. A data set can have one mode (unimodal), two modes (bimodal), or more (multimodal). The way to calculate the mode is :

a. Single Data Mode

Example:

Determine the mode of the quiz scores of each of the following groups!

Group a: 10, 3, 10, 10, 12, 15, 25, 10
Group b: 21, 14, 14, 30, 31, 5, 13, 13, 35
Group c: 40, 12, 6, 8, 7, 10, 10, 40, 6, 5, 9
Group d: 3, 5, 20, 10, 18, 30, 50

Completion:

Group a: the mode is 10.
Group b: there are two modes, 13 and 14.
Group c: the modes are plural: 40, 6, and 10.
Group d: has no mode

b. Group Data Mode

$$M_o = \chi_{mo} + w\left(\frac{b}{a+b}\right)$$

Keterangan:
M_o = Modus
χ_{mo} = Batas bawah kelas modus
w = Lebar kelas modus
a = Frekuensi kelas modus dikurangi frekuensi sesudahnya
b = Frekuensi kelas modus dikurangi frekuensi sebelumnya

Data Location Size

The size of the data location is a way of knowing certain locations from a set of existing data In the size of the data location that most often appears or is used is quartiles, deciles, and percentiles. (Adhiwibowo & Putri, 2023)

1. Quartiles

Quartiles are dividing a stretch of data into four equal parts after the data has been sorted from smallest (x_min) to the largest (x_max)

Single data:

a. Determine the median (Q_2) by dividing the stretch of data into two parts
b. Q_1 (lower quartile) is the median of the left bentagan data
c. Q_2 (upper quartile) is the median of the right-hand stretch of data

$$Q_i = L_{Q_i} + \left(\frac{\frac{i}{4}n - \sum f_i}{f_i}\right)c$$

i = 1, 2, 3
L_{Q_i} = lower edge of quartile class
n = a lot of data
$\sum f_i$ = cumulative frequency before quartile class
f_i = quartile class frequency
c = class length

2. Deciles

Decile is a term that indicates the division of data into 10 equal parts. (Juniardi & Natasha, 2023). Decile values are very likely to be decimals.

The formula for the location of single data deciles is

Group data:

$$D_i = L_{D_i} + \left(\frac{\frac{i}{10} n - \sum f_i}{f_i} \right) c$$

i = 1, 2, 3, ..., 9
L_{D_i} = lower edge of the decile class
n = a lot of data
$\sum f_i$ = cumulative frequency before decile class
f_i = decile class frequency
c = class length

3. Percentiles

If decile is dividing data into 10 equal parts, then Percentile is dividing a set of ordered data into one hundred equal parts.

The formula for percentile location is

$$P_i = \frac{i}{100} \times N$$

i = 1,2, 3, ..., 99

$n =$ a lot of data

Meanwhile, the group data uses the following formula:

$$P_i = L_{P_i} + \left(\frac{\frac{i}{10} n - \sum f_i}{f_i} \right) c$$

i = 1, 2, 3, ..., 99
L_{P_i} = lower edge of the percentile class
n = a lot of data
$\sum f_i$ = cumulative frequency before the percentile class
c = class length

Measures of Data Spread

1. Reach or Range (R)

Range is a statistical measure that quantifies how much difference there is between the maximum value and the minimum value in a data set. It gives an idea of the spread of values or variation in the data. Range is calculated by subtracting the minimum value from the maximum value.

$$R = x_{maks} - x_{min}$$

x_{maks} : Maximum statistics or the largest data

x_{min} : Minimum statistics or smallest data

2. Inter-quartile overlay or range (H)

$$H = Q_3 - Q_1$$

Q_1 : first quartile or lower quartile
Q_3 : third quartile or upper quartile

ISSN: 2615-8019

3. Quartile Deviation or Semi-Quartile Range (Qd)

$$Q_d = \frac{1}{2}(Q_3 - Q_q)$$

Average Deviation (S_r)

The mean deviation, or often called mean deviation, is a measure of how far the average value in a data set is from the mean value. Mean deviation is calculated by taking the difference between each individual value in the data set and the average, and then taking the average of those absolute values.

Single data : $S_r = \frac{\sum |x_i - \bar{x}|}{n}$;

Group data : $S_r = \frac{\sum f_i |x_i - \bar{x}|}{N}$;

4. Standard Deviation or Standard Deviation or Standard Deviation (S)

Standard deviation, also known as standard deviation or standard deviation, is a measure of the spread or variation of data within a set. It measures how far the individual values in a data set are spread out from the mean value. Standard deviation is calculated by taking the square root of the variance.

Single data

Variety or Variance : $S^2 = \frac{\sum (f_i (x_i - \bar{x})^2}{\sum f_i}$

Standard deviation : $S = \sqrt{S^2}$

CONCLUSION

Statistics plays a central role in data-related sciences. The definition of statistics has evolved over time, and can be defined as a collection of numbers or rules related to the collection, processing, analysis, and interpretation of data. Descriptive statistics involves describing, collecting, and presenting data through tables or diagrams. Data presentation can be done through various forms of diagrams, including bar charts, line charts, and pie charts. Tables, as a form of data presentation, can be frequency tables, classification tables, contingency tables, or correlation tables. Measures of data centering, data location, and data spread provide insight into the center, distribution, and variation of values in a data set.

Thus, statistics has a crucial role in understanding, analyzing, and making data-based decisions, especially in today's digital era.

REFERENCES

[1] Adhiwibowo, B., & Putri, A. D. (2023). Pengantar Ilmu Statistika: Untuk Mahasiswa dan Umum. Yogyakarta: Anak Hebat Indonesia.

[2] Juniardi, W., & Natasha, P. (2023, January). Pengertian Desil dan Persentil dengan Cara Menentukannya. Retrieved from Quipper: https://www.quipper.com/id/blog/mapel/matematika/desil-dan-persentil/

[3] Nasution, L. M. (2017). Statistik Deskriptif. Jurnal Hikmah, 49-55.

[4] Nisva, T. M., & Ratnasari, V. (2020). Analisis Regresi Logistik Biner pada Faktor-Faktor yang. INFERENSI, 29-36.

[5] Rochman, F., & Kelana, J. B. (2022). PEMBELAJARAN PEMAHAMAN MENGENAI DIAGRAM. Collase, 795-801.