Proceeding International Seminar on Islamic Studies Vol. 6, No. 1 (2025)
Malaysia, February 1, 2025 E-ISSN: 2722-7618 || P-ISSN: 2722-7626

FREQUENCY DISTRIBUTION IN DATA ANALYSIS USING THE R

Hevlie Winda Nazry S[1*], Firahmi Rizky[2], Yohanni Syahra[3]

[*1, 2, 3]Universitas Muhammadiyah Sumatera Utara
[*1]email: hevliewindanazry@umsu.ac.id
[2]email: firahmirizky@umsu.ac.id
[3]email: yohannisyahra@umsu.ac.id

Abstract: Frequency distribution is a statistical technique used to describe the distribution of data in tabular or graphical form. This concept plays an important role in data analysis because it allows researchers to understand data patterns and trends. Frequency distribution can be divided into two types, namely qualitative frequency distribution and quantitative frequency distribution. This research discusses the basic concepts of frequency distribution, the types of distribution, and their applications in various fields such as economics, health, and social. The results show that frequency distributions are effective in identifying data patterns and trends, thus helping to make informed decisions. This research uses the R application which is used to help calculate and visualize data into graphical form.

Keywords: Frequency Distribution, Statistics, Data Analysis, R Program

Introduction

In the modern era dominated by the development of information technology, data analysis has become one of the indispensable skills in various fields. One method that is often used in data analysis is frequency distribution, which is a technique to summarize data in tabular or graphical form to facilitate understanding of data patterns and characteristics. Frequency distribution allows researchers or practitioners to identify trends, outliers, and data distribution more effectively.

R, as one of the most popular open-source software among researchers and data scientists, offers various tools and packages for performing data analysis, including frequency distributions. R is known for its ability to process data efficiently, produce informative visualizations, and support various statistical methods. The use of R not only helps simplify the analysis process, but also allows users to develop more flexible and specific approaches as needed.

This research aims to discuss the concept of frequency distribution and its implementation in data analysis using R. By understanding how frequency distribution is implemented using R, it is hoped that readers can develop more in-depth data analysis skills and make the most of R's potential. In addition, this research will also explain the advantages of R in data analysis and provide practical guidance for producing accurate and informative frequency distributions.

Through this article, it is hoped that readers will not only gain a theoretical understanding of frequency distributions but also have the practical skills to implement them in various cases of data analysis using the R application.

Research Objectives

Frequency Distribution

Frequency distribution is a very important technique in data analysis as it allows the presentation of data in a more structured and understandable form. In data analysis, frequency distributions are used to group data into specific intervals or categories so that distribution patterns, data concentration, and outliers can be clearly identified.

For example, in survey analysis, frequency distributions can be used to describe the number of respondents by age group, education level, or income. Data that has been summarized in the form of a frequency distribution table or graph provides an overview of the characteristics of the population under study.

With the help of the R application, this process becomes easier and more efficient. R provides various functions and packages such as "table(), hist(), and ggplot2" that allow users to quickly create frequency distribution tables and graphs. In addition, packages such as "dplyr" allow data cleaning and transformation before analysis, while "tidyr" helps tidy up unstructured data. Thus, frequency distributions in data analysis become more organized, accurate, and informative when performed using R.

Group Data

From a survey involving 90 students with a range of scores on math scores. The raw data is then grouped into several intervals, as follows:

Grade	f
30-39	5
40-49	10
50-59	15
60-69	25
70-79	20
80-89	10
90-99	5

The frequency distribution above can help illustrate math scores by the number of students in each group. From the table, it can be concluded that the highest frequency of students is in the 60-69 score range.

Methods

1. Data Collection
   Selected Teaching Materials for Probability and Statistics of Data Science Students
2. Data Analysis
   a. Introduction and Frequency Distribution approach to statistical techniques for describing the distribution of data in tabular and graphical form.
   b. Application of R to Frequency Distribution to help calculate and visualize data.
3. Limitation:
   a. This research only focuses on understanding the theory of Frequency Distribution and its application to R
   b. The data used is limited and also relatively simple so that it is easy to understand.

Program R

The frequency distribution can be calculated and visualized using R. Here is a simple R code example to create a frequency distribution table and graph:

```
# Respondent data by age group
data <- c(5, 10, 15, 25, 20, 10,5)
age_group <- c("30-39", "40-49", "50-59", "60-69", "70-79", "80-89",90-99")

# Creating a frequency distribution table
frequency table <- data.frame(Group_Age = group_age, Number_Respondents = data)
print(table_frequency)

# Visualization with bar charts
library(ggplot2)
ggplot(table_frequency, aes(x = Age_Group, y = Number_Respondents)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title="Distribution of Respondents by Age Group",
       x = "Age Group",
       y = "Number of Respondents") +
  theme_minimal()

#Create a histogram
barplot (frequency,
         names.arg = class,
         main = "Histogram of Statistical Test Results",
         xlab = "Value Class",
         ylab = "Frequency",
         col = "lightblue",
         border = "blue")
```

Results and Discussion

In this section, the concept of frequency distribution will be explained, starting with the meaning of frequency distribution, terms in the frequency distribution table, steps for preparing frequency distribution tables, types of frequency distribution tables.

Frequency Distribution

The word *distribution* comes from the word distribution (English), which means channeling, sharing, or radiating. So, basically, frequency distribution can be interpreted as frequency distribution, frequency division, or frequency emission. Meanwhile, *frequency* itself also comes from the English word frequency, which means frequency, frequency, or infrequency. In statistics, frequency means how many times a variable symbolized by a number (number) repeats itself in a series of numerical data.

Thus, frequency distribution is a situation that describes how the frequency of symptoms or variables symbolized by numbers has been distributed, divided, spread, and radiated. The depiction of numbers (numbers) or the presentation of numerical data can be presented in the form of tables or graphs/images, which are then known as frequency distribution tables and frequency distribution graphs. Frequency distribution is the arrangement of data according to certain interval classes or according to certain categories in a list (Hasan 2001 in 2014 Literature review).

A frequency distribution is a series of numerical data according to their quantity and/or quality (categories). A series of numerical data according to its quantity is called a quantitative frequency distribution, whereas data organized according to its quality (categories) is called a qualitative frequency distribution. A simple example of quantitative data is data that includes learning outcomes, learning achievements, the number of students and so on. While examples of qualitative data are data about gender, type of work, education level, marital status and so on. A table is a statistical data presentation tool in the form of rows and columns, thus, the Frequency Distribution Table can be interpreted as a statistical data presentation tool in the form of columns and lanes in which numbers are contained that can describe the frequency distribution of the variables that are the object of research.

In a frequency distribution table, the following terms are recognized:

a. Interval Class
   Interval Class is an interval containing a class. Usually, a frequency distribution table consists of several classes, generally 5 (five) to 15 (fifteen) classes.
b. End of class
   The end of the class is the value of the end of each class in the distribution that is used as a guideline to put the observed numbers into the class. There are two ends in each class, namely the lower end of the class and the upper end of the class.
c. Class limit
   The class limit is the number obtained by subtracting or adding the end-of-class values with the level of accuracy used. In this case, the level of accuracy of the data used depends on the data recording. If the data used is written in whole numbers, then the level of accuracy of the data is 0.5. if the data used is written in one decimal number, then the level of accuracy of the data is 0.05, two decimal numbers the level of accuracy of the data is 0.005 and so on. There are two class limits in each class interval, namely the lower limit of the class and the upper limit of the interval class. The numbers for the lower limit of the class are obtained from the values of the lower end of the class minus the precision of the data, while for the upper limit of the class plus the precision of the data.
d. Center value
   The middle value is the calculated average of the two ends of the class. How to calculate it: $Nilaitengah$ = 1 2 ($ujungbawa\ h\ kelas$ + $ujungataskelas$) e. Class length Class length is the number obtained from the difference between the two class boundaries.

Compiling Frequency Distribution Tables

To compile a frequency distribution table with the same class length, the following steps are taken:

a. Range (R).

$$Rentang\ (R)\ =\ Nilai\ Maksimum\ -\ Nilai\ Minimum$$

b. Determining the Number of Classes (B). The number of classes used is at least 5 classes and at most 15 classes, Another way for large n ≥ 100 is to use Sturges' rule, namely:

$$Banyak\ kelas\ (B)\ =\ 1\ +\ 3{,}3log\ n$$

c. Determine the class length (P), which is the quotient of the Range with the Number of Classes using the formula:

$$Panjang\ Kelas = \frac{Rentang}{Kelas}$$

d. Determine the ends of the class for each interval class. In determining the ends of the class that must be considered is determining the lower end value for the first interval class. there are two possibilities that can be done, namely: the lower end value of the first interval class can take the smallest data value or a data value smaller than the smallest data value.
e. Entering all data into each interval class, using the tally column.
f. Write the number and title of the table as well as the description and source of the data obtained.

Frequency Distribution of Qualitative Data

A frequency distribution of qualitative data is a collection of predefined values and their frequencies. The frequency distribution of qualitative data is useful for providing a table of observed values and showing how often something occurs. The frequency distribution of qualitative data is divided into two namely; Relative Frequency Distribution and Percentage Qualitative data.

The relative frequency of a class is the proportion of items in each class number to the total number of items in the data. If a group of data has n observations, then the relative frequency of each category or class will be given as follows:

Relative frequency of a class $\frac{f}{n}$

Where:
f : Class or group frequency
n: total frequency value
While the percentage frequency of a class is the relative frequency of the class multiplied by 100.

Frequency Distribution of Qualitative Data

There are three things that need to be considered in determining the classes for the frequency distribution for quantitative data, namely the number of classes, class width and class boundaries. After conducting a trial on the system, it can be concluded that the results obtained are:

1. Number of Classes
   With the following formula:
   K = 1 + 3.322 log n
   Where k = number of classes
   N = number of observation values
2. Class Interval
   To determine the class size (interval length), the formula was used:
   $$c = \frac{Xn - X1}{k}$$
   where: c = estimated class size (class width, class size, class length)
   k = number of classes
   Xn = largest observed value
   X_1 = smallest observed value
3. Class limit
   The distance between the upper class limit and the lower class limit is also called the class width or length. The lower class limit indicates the smallest possible data value in a class. While the upper class limit indicates the largest possible data value in a class.

Data

Grade	f
30-39	5
40-49	10
50-59	15
60-69	25
70-79	20
80-89	10
90-99	5

Table 1. Student score data

1. The relative frequency value as well as the percentage value of the data above and the range of the data above along with the interval class.
2. Draw a histogram

The answer:

1. a). Relative frequency values

$= F = \frac{f}{n}$

GRADE VALUE	Relative frequency	Frequency percentage
30-39	0,05	5
40-49	0,1	11
50-59	0,16	16
60-69	0,27	27
70-79	0,22	22
80-89	0,11	11
90-99	0,05	5

Table 2. Frequency Distribution

F30 – 39 = $\frac{5}{90}$ = 0.05

F 40 - 49 = $\frac{10}{90}$ = 0.11

F 50 - 59 = $\frac{15}{90}$ = 0.16

F 60 - 69 = $\frac{25}{90}$ = 0.27

F 70 - 79 = $\frac{20}{90}$ = 0.22

F 80 - 89 = $\frac{10}{90}$ = 0.11

F 90 - 99 = $\frac{5}{90}$ = 0.05

b). Fk x 100

= 0,05 x 100 = 5

= 0,11 x 100 = 11

= 0,16 x 100 = 16

= 0,27 x 100 = 27

= 0,22 x 100 = 22

= 0,11 x 100 = 11

= 0,05 x 100 = 5

2. Range = R = Xman - Xmin
   R = 99 - 30 = 69
   K = 1 + (3.3) log n
   K = 1 + (3.3) log 69
   K = 1 + (3,3) 1,838
   K = 7,065
   K = 7 (interval class)
   Class length: P = $\frac{R}{K}$
   = P=$\frac{69}{7}$ = 9.85
   = P= 9 or 10

3. Histogram:

Image1 . Results of R Application

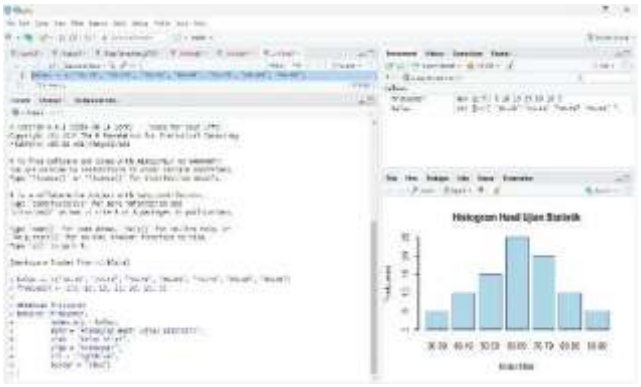

Image2 . Frequency Distribution of Student Data On Program R

```
#Data
class <- c("30-39", "40-49", "50-59", "60-69", "70-79", "80-89", "90-99")
frequency <- c(5, 10, 15, 25, 20, 10, 5)
#Create a histogram
```

```
barplot (frequency,
    names.arg = class,
    main = "Histogram of Statistical Test Results",
    xlab = "Value Class",
    ylab = "Frequency",
    col = "lightblue",
    border = "blue")
```

Coding Create Histogram for image 2

Conclusion

Frequency distribution is a statistical technique used to describe the distribution of data in tabular or graphical form. In this writing, we have found several formulas used in Frequency Distribution, namely the formula for finding relative frequency $=\frac{f}{n}$, the formula for finding quantitative frequency, namely by multiplying the result of the relative frequency by one hundred, the formula for finding the range, number of classes and also the class length. From the results of this study that Frequency Distribution can help in identifying patterns and outliers in the data. Not only that, Frequency Distribution analysis can help choose the right statistical model in processing the data. From this problem, it can be seen that the relative frequency can be determined if the data has a class and also a frequency. Then, if the relative frequency has been found then the quantitative frequency can be generated. In finding the class length of the data, first find the range and also the interval class. So when it is further examined in solving or processing data using Frequency Distribution, every problem solving is still continuous with the previous formula.

And by using the R application, calculations and data visualization can be done easily and efficiently. Where the R application provides several features that can support this Frequency Distribution analysis. Such as by providing table(), hist(), and plot() formats so that when the user wants to visualize the data to be analyzed, the user just types the command in the script column available in the R application.

Bibliography

Wahab, A. 2021. Presentation of Data in Frequency Distribution Tables and Applications in Education Science

Sugiyono. 2009. Quantitative Research Methods Qualitative and R & B. Bandung

Sudijono, A. 2018. Introduction to Educational Statistics. Depok

Sudjana. 1996. Statistical Methods. 6th Edition. Tarsito: Bandung

Pratama, R. & Yuliana, M**.** (2019). Application of frequency distribution for data analysis with R software. Journal of Statistics and Data, 18(3), 112-120.

Kusuma, S., & Naufal, H. (2021). Frequency distribution analysis using R for numerical data. Journal of Engineering and Statistics, 14(4), 89-98.

Suwarto, A. & Setiawan, E**.** (2020). Frequency distribution in data analysis using the R application: Implementation and application in statistical research. Journal of Computer Science and Statistics, 15(2), 234-245.

Lestari, P., & Aulia, N**.** (2018). Statistical programming with R for frequency distribution analysis on big data. Journal of Mathematics and Statistics, 22(1), 45-59.

Yuliana, M., & Lestari, P. (2021). Using R to analyze frequency distribution on big data. *Journal of Computing and Statistics*, 23(1), 56-70.

Sari, N., & Kusuma, D. (2022). Frequency distribution analysis using R application in scientific research. *Journal of Applied Statistics*, 10(3), 150-161.

Setiawan, B., & Rinaldi, H. (2018). Frequency distribution application for data analysis with R: Theory and implementation. *Journal of Technology and Statistics*, 17(1), 22-34.

Budianto, H., & Rinaldi, S. (2014). Frequency distribution programming in R for statistical data analysis. *Journal of Mathematics and Statistics*, 11(2), 102-113.