Journal of Marketing and Consumer Research
ISSN 2422-8451 An International Peer-reviewed Journal
Vol.90, 2023

Business Research Methods: Determinants of Usage of Graphic Presentation and Statistical Tests in Business Research

Austin Mwange, Ph.D./DBA*
Senior Lecturer, Department of Economics, School of Social Sciences, ZCAS University, Lusaka, Zambia; Lecturer Graduate School of Business, University of Zambia, Lusaka, Zambia; Lecturer, University of Lusaka, Lusaka, Zambia
Email: austin.mwange@unza.zm; austin.mwange@zcasu.edu.zm; austin.mwange@unilus.ac.zm austin.mwange@unza.zm

Joseph Chiseyeng'i, Ph.D.
Doctorate in Business Administration Student, Department of Business Administration, School of Business, ZCAS University, Lusaka, Zambia; Senior Lecturer, Business Department & Zambia University College of Technology (ZUCT), Ndola
Email Address: bashikapembe@gmail.com

Windu Matoka, DBA
Senior Lecturer, Department of Business Administration, School of Social Sciences, ZCAS University, Lusaka,Zambia
Email: windumatoka@gmail.com; windu.matoka@zcasu.edu.zm

Sampa David Shiyunga
PhD Economics Student, University of Cape Town, South Africa
Email: shysam001@myuct.ac.za or dshiyunga1@yahoo.com

Abstract
Data summarization and presentation skills, as well as data analysis skills through the application of statistical tests, are important skills that every graduate student must master. These skills enable researchers to reach research conclusions and recommendations. The main purpose of this work is to first describe the best selection of tables and graphs for examining and describing various aspects of the data in the study. Second, we describe the relationships between variables and trends in study data and discuss the selection of the most appropriate statistical tests for examining them. The methodology used in this study was descriptive as it discussed and explained how the data were presented and analyzed. This white paper will therefore consist of two parts: Part 1 focuses on summarizing and presenting data using tables and graphs, while Part 2 focuses on data analysis by applying various statistical tests.
Keywords: Tabular and graphic data presentations; Statistical tests, Data analysis, Parametric Tests, t-Test, Z-test, Chi-Square, ANOVA, Regression Analysis
DOI: 10.7176/JMCR/90-04
Publication date: July 31st 2023

1. INTRODUCTION

Among the many investment options available in Zambia today, pension funds are a popular choice for those looking to invest in their future. If you want to decide which pension fund to invest in, then you want to know which pension fund you can invest in. They also want to know how individual funds have performed in the recent and distant past. Also, we need to look at both qualitative and quantitative variables and finally see a lot of data. The questions you ask yourself are: "How do I summarize this wealth of data and present it in a way that facilitates making the best investment decisions?" One of them would have asked himself. Researchers have to answer during the research process.

Research is the process of starting with research questions and goals, planning methodology, data collection, and finally selecting the most appropriate statistical tests to facilitate data analysis. In most cases, at the final stage of data analysis, researchers are asked to summarize the data they have collected and present it in the form of tables and graphs. Tables and graphs are visual representations used to summarize, organize, and present data in highly meaningful ways to show patterns and relationships. This allows researchers to easily interpret and visually understand research results. Summarizing and graphing data requires the researcher's knowledge of the data's characteristics, the ability to identify appropriate tables or graphs to apply, and proficiency in preparing them.

However, despite the fact that tables and graphs are useful tools for helping researchers make informed decisions, they only provide part of the whole story. Researchers will be required to make inferences about the population from the data collected and this will require performing statistical tests. Statistical tests are analytic techniques that are applied to enable the research to make conclusions on the data. Therefore, not only will the researcher be required to summarize and organize data in tables and graphs, he or she will also be required to perform statistical tests on the organized data to arrive at research conclusions and recommendations. But while tables and graphs are tools that help researchers make informed decisions, they only tell part of the story. Researchers need to draw inferences about the population from the data collected, which requires performing statistical tests. Statistical tests are analytical techniques used to enable studies to draw conclusions about data. Therefore, researchers should not only summarize and organize data in tables and graphs, but also perform statistical tests on the organized data to reach research conclusions and recommendations.

This article discusses different types of tables and graphs available to the researcher for data summarization and presentation. It further discusses the various statistical tests that the research can perform to either confirm or reject research hypotheses. Of necessity, it begins with defining data, its characteristics and classification, before discussing how it can be presented in tables and graphs. It discusses tables and graphs that are commonly used in research including Frequency tables, Histograms, Bar charts and Scatter diagrams. It then proceeds to define Statistical tests and discusses parametric tests such as the T-test, Analysis of Variance (ANOVA), as well as non-parametric tests such as the Chi-squared test.

A presentation of existing knowledge rather than a paper aimed at developing original knowledge, this paper describes various types of tables and diagrams that researchers can use to summarize and present data. increase. It also describes various statistical tests that studies can use to confirm or reject research hypotheses. Inevitably, we start with the definition of data, its properties, and classifications, and then discuss how to represent the data in tables and graphs. It covers tables and graphs commonly used in research, including frequency tables, histograms, bar charts, and scatter plots. It then defines statistical tests and discusses parametric tests such as the t-test, analysis of variance (ANOVA), and nonparametric tests such as the chi-square test.

2. METHODOLOGY

The methodology used in this study was descriptive, as it describes the aggregation and graphing of data and the performance of statistical tests to enable researchers to reach study conclusions. Papers eligible for this study were primarily cited from statistics textbooks and academic papers from Google Scholar, EBSCO, and JSTOR. In my initial search, the search terms were:Tables and graphs used in surveys, statistical tests and statistical data analysis.

3. DISCUSSION

This is a two-part discussion that first analyses data summarisation and presentation before discussing the various statistical tests that can be performed on data thus presented. Part One is entitled; *Tabular and Graphic Presentation of Data in Research,* and Part Two is entitled; *Statistical Tests Applied in Research.*

3.1 TABULAR AND GRAPHIC PRESENTATION OF DATA IN RESEARCH

Definition and Classification of Data

Before discussing the different types of tables and graphs used in research, it's worth defining key concepts such as data, variables, and observations. During fieldwork, researchers gather information through questioning, systematic observation, or laboratory testing. The information thus collected constitutes the data for research and individual data sets permit statistical testing and analysis. Data can be classified as either qualitative or quantitative. Observations are facts, numbers, or measurements recorded for a particular variable. For example, if a researcher observes the variable gender in her sample of 30 students, we can say that the dataset contains 30 observations of her. In tables and graphs, the components being compared or measured are called variables. A variable is a measurable characteristic or attribute that makes up data and can take on many different values. For example, if your research question is:How does a person's weight depend on their height? Height is one variable and weight is another. It is often convenient to describe variables as dependent or independent. The dependent variable is that which changes relative to each level of the independent variable (Upagade, 2012).

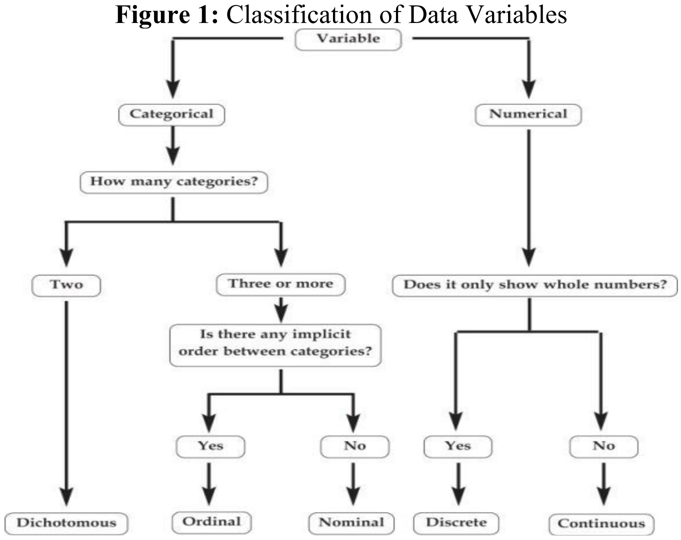
Qualitative or Categorical Data

Variables can be broadly divided into two groups. Both qualitatively and quantitatively. A qualitative (or categorical) variable is a category of variables that contain non-numeric data attributes such as gender, nationality, blood type, and so on. Qualitative data can be further decomposed into dichotomous, nominal and ordinal variables. A dichotomous variable, also called a binary variable, is a variable that has only two categories. So there are only two possible answers (male or female, yes or no). An ordinal variable is a variable with three or more categories that can be ordered ascending or descending. A nominal variable is a variable with three or more categories without a definite order of the categories, such as: B. Blood types A, B, AB, and O (Murphy, 2019).

Qualitative or Numerical Data

.Qualitative or numeric variables are variables that can be counted and are further divided into continuous and discrete variables. A discrete variable is an observation that can only take on certain numeric values, such as: B. Number of students in class. A continuous variable is a variable that is measured on a continuous scale and has as many decimal places as the meter can capture. These include blood pressure, birth weight and height. This data classification is shown in Figure 1.

Figure 1: Classification of Data Variables



In research, discrete or continuous variables measured on a numerical scale are richer in information and are preferred for parametric statistical analysis and tests.

Forms of Data Presentation

Before data is collected and analyzed, it should be categorized and organized in a way that is easy to read and interpret. In other words, data must be transformed into information to support informed decision-making. Data is usually large in size and needs to be transformed into a compact, easy to see and read format by displaying it in tables, charts, graphs or charts. This is consistent with the words of an unknown author. Therefore, researchers must learn to transform data into information through graphical representations. In general, he has three forms of data presentation in research papers. Textual or descriptive presentation, tabular presentation, schematic or graphic presentation (Murphy, 2019).

In-textual Presentation of Data

In-textual presentation is when data is described within the text. This form of presentation is more suitable when the quantity of data is not too large. Consider the following example:

The 2015 Living Conditions Monitoring Survey (LCMS) results show that the population of Zambia was estimated at 15.5 million in 2015. This population was mainly concentrated in rural areas at 58.2 percent compared to 41.8 percent in urban areas. About 1,700, 000 households were in rural areas while about 1, 300,000 were in urban areas (LCMS, 2015).

This is an example of data-in-text presentation that has the advantage of allowing researchers to emphasize specific details within the presentation. A serious drawback of this method of presentation is that the reader must read the entire presentation text to understand it. A general rule for using representations in text is that the data presented in figures and tables should not be repetitions of data already contained in the text (Flockton, 2004).

Tabular Presentation of Data

In tabular presentation, data are presented in rows and columns. The most important advantage of tabulation is that it organizes data for further statistical treatment and analysis. Classifications used in tabulation are of four kinds:

1. *Qualitative* - when classification is done according to attributes, such as social status, physical status, nationality
2. *Quantitative* - when the data are classified on the basis of characteristics which are quantitative in nature; that is characteristics can be measured quantitatively such as age, height, production, income
3. *Temporal* – when time becomes the classifying variable and data are categorised according to time
4. *Spatial* - when the data is classified on the basis of place; a village, town, or country

Diagrammatic and Graphic Presentation of Data

This is the third data presentation method after the text and tabular methods. A chart or heart displays information by representing it as a shape. Compared to textual and tabular representations, this method provides the fastest understanding of the actual situation described by the data. Graphical representations of data are very effective in translating the very abstract concepts contained in numbers into a more concrete and understandable form. Charts are less precise, but they are much more effective at presenting data than tables (Duquia, 2014). Various types of charts are available to researchers. The most important of these are frequency charts, histograms, line charts, bar charts, pie charts, and scatter charts. A researcher's rationale for choosing the type of table, chart, or graph to use depends on what the results are trying to explain. For example, a scatterplot is used to show the relationship between her two variables, and a pie chart is used to show the relative contribution of an element to the whole. Charts, tables, and graphs are valuable because they help researchers identify patterns and trends that may be less obvious in statistical results. It also helps researchers understand which direction to take statistical tests.

Tables, Charts, and Graphs Used in Research

The Choice of Tables and Graphs

Data can be represented graphically as text, tables, or charts, graphs, or graphs. Each of these may be suitable for providing information to the reader. Whenever possible, researchers should use representations of data in text. However, if the researcher cannot present the data in her sentence or two, a table or graph is appropriate. The overall goal is to enable researchers to use this method to convey the meaning of their data to readers in an informative, interesting and engaging way. The type of table or chart you use depends on the main purpose and target audience of your presentation. Tables are generally better than charts for providing structured numerical information, but charts are better for showing trends, making broad comparisons, and showing relationships. (Murphy, 2019). When choosing a presentation format, researchers should ask themselves the following basic questions:what am I trying to tell you? Who is my audience? What type of presentation works best? Which numbers are easiest for readers to understand? What will help them better understand the data? What is the easiest and most effective way to present the data?

Shende (2010) advises that when considering the characteristics of good tables and graphs, it is worth emphasizing that tables or graphs:

- *Should be self-explanatory* - readers should be able to get information from the tables or graphs without reading the whole text. On the other hand, if warranted, the text should always include mention of the key points in a table or figure
- *Should have a clear descriptive title* - which will serve as the *topic sentence* of the table or graph. The titles can be lengthy or short, depending on the discipline
- *Should have column titles* - the goal of which is to simplify the table. The reader's attention moves from the title to the column title sequentially. A good set of column titles will allow the reader to quickly grasp what the table or graph is about
- *Should have a table or graph body* - This is the main area of the table or graph where numerical or textual data is located. Tables and graphs must be constructed so that elements read from up to down, and not across
- *Should provide additional information* - in the key or footer, further information about the table or graph footer must be provided if needed.
- *Should be properly referenced* – table or graph should be inserted into a document only after being mentioned in the text. They must be properly referenced and quote the source which provided the data, if required. It must be ordered in the number that it appears in the text. In addition, tables and graphs should be set apart from the text
- *Should be centered* - the placement of figures and tables should be at the center of the page
- *Should not have text wrapping* - sometimes, tables and figures are presented after the references in selected journals

Presentation of Data

We now turn our focus to how data can be summarized and presented in tabular or graphic forms. We will first consider tabular presentation of data, followed by diagrammatic or graphic presentation.

Tables

A table is a visual representation that organizes information into rows and columns that facilitate comparing and contrasting information. Tables are also great for showing patterns and relationships. Figure 2 below shows Zambia's population growth over the past decade.

Figure 2: Zambia Historical Population Growth Rate Data

Year	Population Growth Rate	Growth Rate
2022	20,017,675	2.80%
2021	19,473,125	2.88%
2020	18,927,715	2.98%
2019	18,380,477	3.05%
2018	17,835,893	3.11%
2017	17,298,054	3.16%
2016	16,767,761	3.20%
2015	16,248,230	3.24%
2014	15,737,793	3.30%
2013	15,234,976	3.33%

Source: https://www.macrotrends.net/countries/ZMB/zambia/population-growth-rate

From this table, it is easy to see that Zambia's average population growth rate over the past decade has been 3% per annum. Researchers have found that tabular data representations are useful for summarizing and simplifying large, complex data. Also, use tables to show trends and patterns across data sets. Most importantly, aggregating data is the basis for creating graphs and charts and forms the basis for statistical analysis and advanced data analysis.

Data aggregation has drawbacks that limit its use. Data types affect the use of tables. For example, tables work better for integers than approximations. Because of their structure, tables take longer to read and understand. Also, they may not provide an easy way to see embedded trends in the data. A tabular representation presents the data in an aggregated form and may not focus on individual elements. The tabular representation shows only numbers and attributes of these numbers cannot be tabulated. Tables may require expertise for the reader to decipher the message buried deep within the tabular presentation. This means that interpreting the table may require the necessary expertise. However, despite its limitations, tabular data representation is the most powerful and versatile data representation method and serves as the foundation for other methods. This is the data presented in the table that serves as input for statistical testing (Murphy, 2019).

Frequency Table

A frequency table, also called a frequency distribution, is one of the most basic tools for displaying descriptive statistics and is used to describe the number of occurrences of a particular variable in a data set. Frequency tables display large amounts of data in a fairly concise manner, are easy to interpret, and are often used as a quick reference for distributing data. They help readers understand obvious trends and can be used for comparisons between datasets. In a particular survey conducted in Zambia, Figure 3 below is a frequency chart of respondents' views on why men continue to defile children despite severe punishments for convicted molesters.

Figure 3: Frequency Table of Reasons for Defilement (Per hundred of respondents)

Response	Frequency
For Rituals	59
Ignorance	18
Sex Boosters	2
Drug Abuse	9
Deliberate/Intentional	10
Cure for HIV and AIDs	2

Source: Nkandela, 2001

It is easy to see that the biggest cases of corruption involve spiritual rituals. One advantage of frequency tables is that they help researchers examine the relative frequency of each specific target variable in a dataset. However, frequency tables are not suitable for all applications. The disadvantage is that complex data sets presented in the form of frequency tables are difficult to understand. The skewness and kurtosis of frequency table data are not readily apparent if they are not displayed in a histogram (Murphy, 2019). Therefore, frequency tables are not suitable for analyzing the distribution of data. Data in the form of clustered frequency distributions are usually further displayed in the form of histograms, frequency polygons, frequency curves and ogives.

Histogram

Of the several ways to graph frequency distributions, histograms are the most popular and widely used in practice. A histogram, used to visually represent data, is a series of vertical bars whose area is proportional to the frequency of the classes they represent. Histograms are primarily used to display data for continuous variables, but can also be adjusted to display discrete data by applying an appropriate continuity correction. A histogram is created from the frequency table as shown in Figure 4 below.

Figure 4: Grouped Frequency Distribution of Research Methodology II Assignment 2 Scores

Interval's Lower Limit	Interval's Upper Limit	Class Frequency
39.5	49.5	3
49.5	59.5	10
59.5	69.5	53
69.5	79.5	107
79.5	89.5	147
89.5	99.5	130
99.5	109.5	78
109.5	119.5	59
119.5	129.5	36
129.5	139.5	11
139.5	149.5	6
149.5	159.5	1
159.5	169.5	1

From the above Frequency table is created the following Histogram.

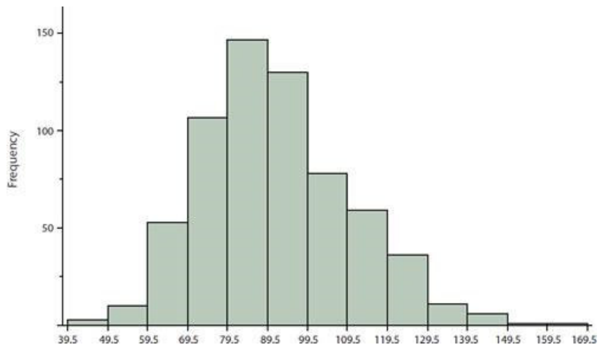


Figure 5: Research Methodology II Assignment 2 Scores Histogram

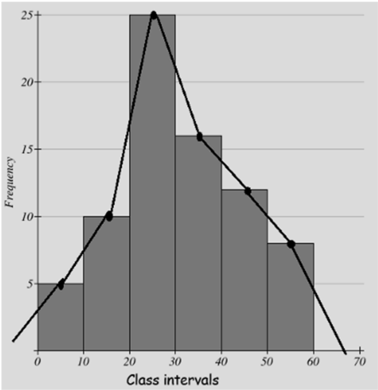
This histogram shows that most values are in the middle of the distribution with few extreme values. The distribution is skewed and not symmetrical because the values spread more to the right than to the left. If the histogram bars take the shape of a bell curve, you can assume that the mean and median of the data are approximately equal and that the data are normally distributed.

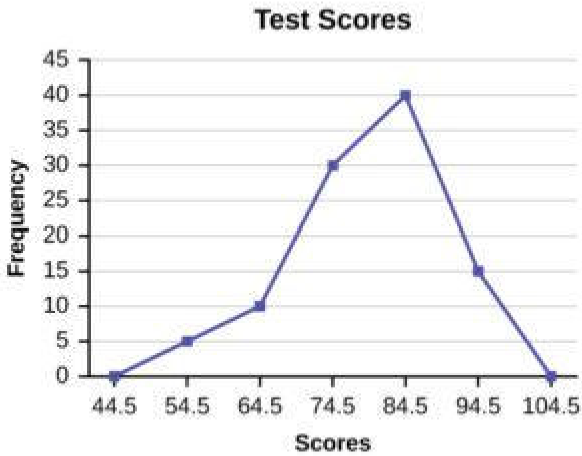
Using histograms has advantages for researchers. They are also used to draw frequency polygons, allowing researchers to get an idea of the shape and spread of the data with a quick glance at the histograms. One of the drawbacks of histograms is that they can only be used for quantitative values. In order to work with qualitative data, surveys need to draw bar charts. It cannot be used to represent discrete frequency distributions, only continuous frequency distributions. Histograms allow researchers to see the shape of the distribution, but they do not provide information about the median, upper quartile, and lower quartile of the data. Using histograms he cannot compare two different data sets.

Frequency Polygons

A frequency polygon is an alternative frequency distribution plot to a histogram. This is the most common way to represent clustered frequency distributions. In the frequency polygon, frequencies are plotted against the midpoints of the class intervals, carefully connected by a series of short straight lines. Figure 6 below is an example of a frequency polygon.

Figure 6: Research Methodology II Assignment 2 Scores Frequency Polygon





Frequency polygons not only help researchers sort and plot data, but also facilitate comparison and contrasting of results. They are very easy to understand and give a clear picture of the data distribution.

Ogives

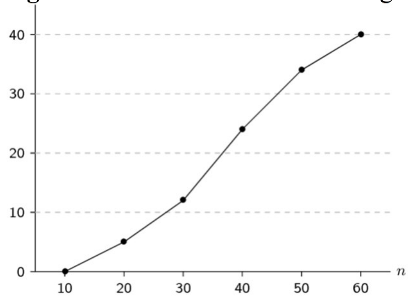
An ogive is a graphical representation of the cumulative frequency distribution of a continuous variable. It consists of plotting cumulative frequencies along the y-axis against class breaks along the x-axis. Therefore, there are two types of ogives for cumulative frequencies: less than and cumulative frequencies. Ogies are used to visually represent sums above or below a certain upper class limit. Provides an overview of each interval in the frequency distribution, showing the percentage of data points above or below a certain value. It distinguishes rates of change between classes better than other charts. The relative slope between points indicates the magnitude of the increase. A steep incline indicates a steeper climb compared to a gentler incline. All counts have already been added to the previous sum, so the final value of the cumulative count is equal to the total number of data values. Ogie is built from a cumulative frequency table like this:

Table 2: Cumulative Frequency of Class Scores

Class Interval	Frequency	Cumulative frequency
10 - 20	5	5
21 - 30	7	12
31 - 40	12	24
41 - 50	10	34
51 - 60	6	40

Computing all the coordinates and connecting them with straight lines gives the following Ogive.

Figure 7: Class Score *Less Than* Ogive

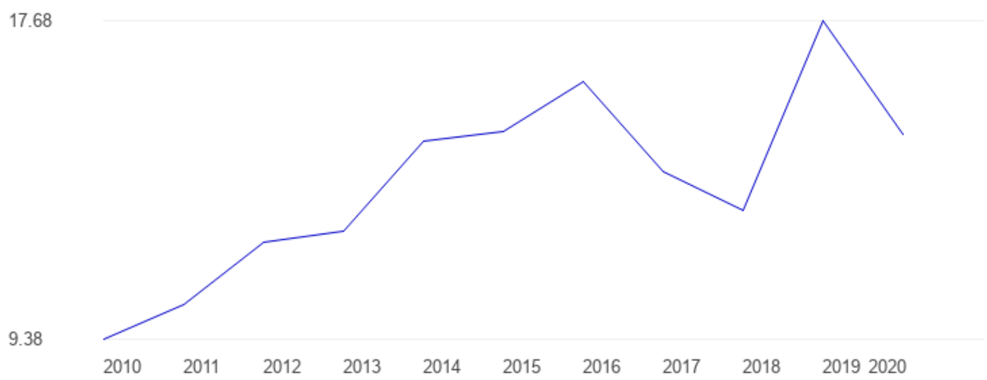


Ogives is especially useful for graphically calculating median, quartile, decile, and percentile cutoff values. It can also be used to graphically determine the number or proportion of observations. An interesting feature of the two ogives together is that their intersection gives the median. Despite its advantages, OGIVE is very expensive to create. Also, the x-axis only plots class boundaries, so it does not reflect all data points. Ogien doesn't reveal much about central tendency and variance. It also doesn't reveal much about the skewness and kurtosis of the data.

Time Series or Histogram

A time series, also called a histogram, is a series of numerical values that show the change of a variable over time. The horizontal axis is labeled as the time axis, but the y-coordinate is the value taken by the variable at a specific time. Similar to a line chart, consecutive points are connected by straight lines. A histogram is a very sophisticated means of predicting the value of a variable under the assumption that history repeats, because it shows the trend and behavior of the variable. Zambia's government expenditure as a percentage of GDP is shown in the histogram below.

Figure 8: Zambia Government Spending as Percent of GDP



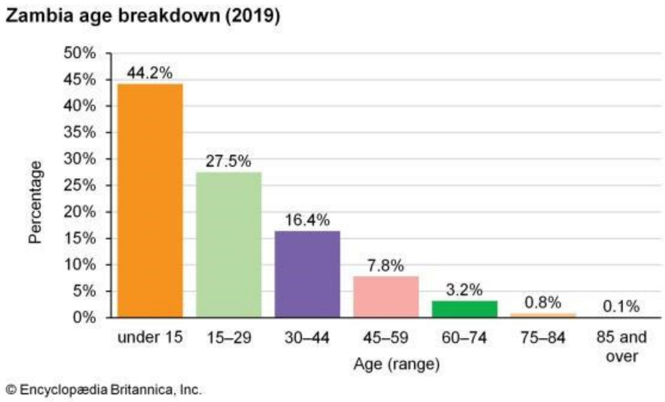
Source: TheGlobalEconomy.com

As can easily be discerned from this Histogram, the Zambian government spending as a percentage of its Gross Development Product has steadily been increasing over the years.

Time series analysis has many advantages for researchers. It is especially useful for predicting future data points. It also helps clean or equalize data, allowing researchers to filter out noise and find true signals in data sets. A major advantage of time series analysis is that it reveals patterns in the data that can be used to predict future data points. It is this predictive aspect that makes time series analysis so popular in research. A disadvantage of time series is that past trends are assumed to continue indefinitely, and valid conclusions can be drawn by extrapolating data based on historical information. In practice, dependent factors can be influenced by a variety of factors, some of which may be unpredictable (Murphy, 2019).

Bar charts

Bar charts are one of the most popular ways to visually display data, and their primary purpose is to display quantities in the form of bars. It consists of a series of bars whose height is proportional to the frequency they represent. Histograms should be clearly distinguishable from bar charts. The most notable physical difference between these two charts is that unlike bar charts, there are no gaps between successive rectangles in the histogram. Histograms are used to show how the data are distributed for each variable, while bar charts are used to show how often a variable occurs. For example:

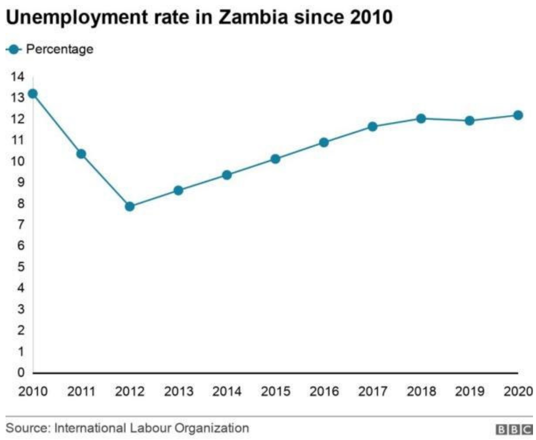


As can easily be discerned from the table, Zambia has a huge youth population; as at 2019 the percentage of Zambian below the age of 30 years was 72%.

One advantage of using bar charts is that you can easily read the quantity based on the height of the bars. Bar charts are useful for researchers because they visually summarize large data sets by displaying each category of data in a frequency distribution. Valuable comparisons are possible and can be used for non-numerical data. It makes trends clearer than tables and estimates key values at a glance. A general drawback is that it cannot be used for continuous variables and requires additional explanation.

Line graphs

Line graphs are usually intended to show the frequency of different values of a variable. Consecutive points are connected by line segments so that the reader can understand the distribution of the variables just by looking at the graph. The simplest line chart is a single line chart, so called because it shows information about the frequency of a single variable. Figure 9 below is a line chart of Zambia's unemployment rate since 2010.



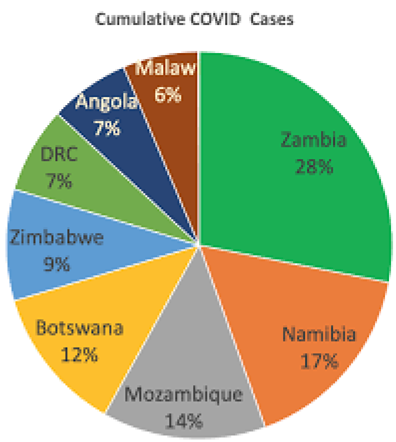
The Line graph clearly shows that Zambia's unemployment rate was between 2010 and 2012, but has been steadily rising ever since.

A line graph can also be converted to a time series to graphically represent information that changes continuously over time. This helps identify trends and allows researchers to make predictions about data that has not yet been recorded. However, line charts do not tell us anything about the cause behind the fluctuations in the data. It also doesn't tell us much about the skewness or kurtosis of the data. The technique of plotting two or more line graphs on the same graph facilitates comparisons between related phenomena (Upagade, 2012).

Pie Charts

A pie chart or pie chart represents a distribution of relative frequencies or percentages, where the circle is divided into sectors with areas proportional to the values of the relative frequencies or percentages. It's probably the simplest chart you can use to display your data, and that's another reason why it's so limited in how it's displayed. However, a quick pictorial of relative class sizes shows when one class is more important than another. Can be used to compare the same item, but two or more populations are different. This can be easily seen in Figure 9 below.

Figure 9: Cumulative COVID Cases in Southern Africa



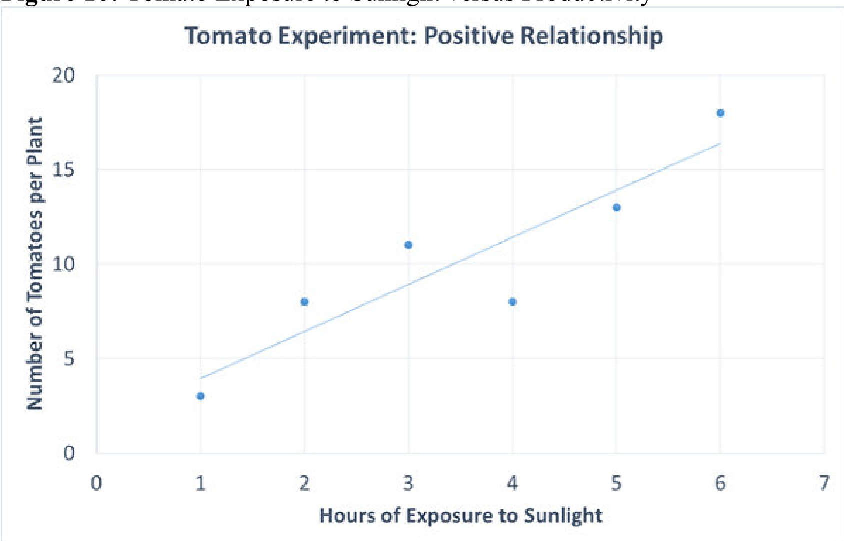
This graphic representation reveals that in the Zambian neighbourhood, one in every three COVID cases is Zambian.

Pie charts have the advantage of being visually appealing and easy to understand. Compared to bar charts and histograms, they are easier to read for those who are not statistically trained. Pie charts allow researchers to easily make broad comparisons across different categories. However, when the series is divided into many components, or when the differences between components are very small, pie charts are less useful for accurate reading and interpretation than bar charts (Murphy, 2019). Similarly, if most data slices are roughly the same size, researchers cannot make visual comparisons between categories just by looking at pie charts. Pie charts cannot be used to represent time series data or to compare two data sets. Pie charts are less popular among professional statisticians and researchers because it is difficult to compare circles.

Scatter Diagrams

A scatter diagram concludes the description of the different types of tables and graphs used in research. Scatter plots are used to show the results of correlation tests aimed at determining whether there is a relationship or association between two variables. If you suspect a causal relationship between two variables, a scatterplot can prove the relationship. A study investigating the relationship between the amount of sun exposure and the number of tomatoes on a plant produced the following scatterplot:

Figure 10: Tomato Exposure to Sunlight versus Productivity



This reveals that there is a positive correlation between the number of hours a plant is exposed to sunlight, and its tomato productivity.

Scatter diagrams are very easy to draw, easy to understand and interpret. An uptrend rising from the lower left corner and moving to the upper right corner indicates a positive correlation. A downward trend from the upper left corner to the lower right corner indicates a negative correlation. Scatterplot flaws show direction, but not correlation or cause. Researchers need to calculate the correlation coefficient to get an accurate value for the degree of correlation.

3.2. STATISTICAL TESTS USED IN RESEARCH

Now that we have discussed the mechanisms for data aggregation and presentation, we will now discuss the statistical tests available to assist researchers in analyzing data and drawing conclusions for their studies. Conducting a survey involves a series of tasks, including planning the survey, designing sample groups, and collecting data. However, data is meaningless unless it is organized and analyzed to generate information for decision making. Researchers are therefore interested in the analysis and interpretation of the data leading to the conclusions of the study. Statistics is all about interpreting data, which is facilitated by performing statistical tests described as a formal procedure for examining researchers' ideas about phenomena (Kothari, 2012). Statistical tests provide a mechanism for making quantitative decisions and seek to determine whether there is sufficient evidence to support or reject a hypothesis. The data analysis aspect of research is particularly challenging, requiring researchers to identify appropriate statistical tests to perform on their data. Luckily, there are a variety of statistical tests to help researchers analyze and interpret data.

The choice of Statistical Tests

A statistical tool used in a statistical test is not an end in itself. As with any tool, the value of statistical analysis tools lies in proper planning and use. Therefore, it is important for researchers to think about the statistical tests and procedures to be used prior to data collection. Choosing the right statistical test depends on the type of variables defined, the variables being compared, and the question the researcher wants to answer.

Researchers should answer the following questions in order to select a statistical test to apply to a particular study. First, what do researchers want to know? Statistical testing and statistical analysis should be related to the research question that determines the methods researchers should use. Second, what kind of data do researchers work with? The types of data they have are basic. This is because statistical tests and techniques that are suitable for quantitative data may not be suitable for qualitative data. Finally, what assumptions do researchers make about the data? For example, it is often assumed that the underlying distribution of data follows a normal distribution.

Three Fundamental Statistical Assumptions

As mentioned above, researchers need to perform statistical tests based on certain assumptions they make about their data. If one or more assumptions are incorrect, the test is not valid and can lead to erroneous results. There are three basic statistical assumptions on which statistical tests are often based. The first assumption is independence of observations. This means that the observations or variables included in the test are not related to each other. The second assumption is homogeneity of variances. This means that the variance within each group being compared should be similar to the rest of the group variances. The effectiveness of the test is limited when the variance of one group is greater than the others. Finally, normality of the data is assumed. That is, the data follow a normal distribution. If your data are consistent with these three basic statistical assumptions, you can perform parametric tests on them (Kothari, 2012).

However, not all datasets meet the three basic statistical assumptions above. Many processes follow non-normal distributions such as beta, exponential, gamma, Poisson and uniform distributions. Even if the data should follow a normal distribution, but they do not, if the sample size is large enough, the researcher can perform a parametric test. Usually 30+ items. He or she may choose to transform the data using various statistical techniques to make them conform to a normal distribution. If the sample size is small, skewed, or represents a different type of distribution, researchers can perform nonparametric tests. Nonparametric tests make fewer assumptions about the data and are useful when one or more of the three basic statistical assumptions are violated. Nonparametric tests draw conclusions that are not as strong as parametric tests.

Statistical Tests Questions

When researchers perform statistical tests, their choice is based on answering one of four questions: What happened, why did this happen, what is likely to happen, and what should we do?

Descriptive statistics testing is the first step in data analysis to answer questions. What happened? For example, researchers want to know how many children from her two-year-old to her ten-year-old are in early childhood education in Zambia. Descriptive tests provide the first layer of information from the collected data. A diagnostic statistical test is her second step in data analysis, following the researcher's answers to the first question. After overviewing what the data is and what the sample looks like, researchers may want to know why certain things are happening. Following the previous example, he or she may find that there are far fewer children in early childhood education in one district than in another. Could there be something in the data that indicates district differences? Using diagnostic analysis, researchers can go a step further and ask the following questions: why did that happen? Predictive statistical tests answer the question of what could happen. Until researchers know why something happened, they can predict what might happen next. By using what is learned in descriptive and diagnostic statistical tests, researchers can perform predictive tests to look into the future. Finally, the prescriptive statistical test answers the following questions. Researchers have an idea of what is likely to happen and may want to know what the best course of action is. This paradigm allows statistical tests to be He falls into two main categories. Descriptive and inferential statistics tests. We first distinguish between descriptive and inferential statistics, and then discuss the statistical tests that fall into each of these categories (Murphy, 2019).

Descriptive versus Inferential Statistics

There are two main areas of statistics. Descriptive and inferential statistics. As the name suggests, descriptive statistics describe a population by summarizing and organizing the characteristics of a sample of data. Inferential statistics, on the other hand, are used to generalize about a population based on sample data. This allows researchers to confirm or disprove hypotheses and assess whether sample data can be generalized to larger populations. Thus, descriptive statistics allow researchers to describe the sample, while inferential statistics allow them to draw conclusions based on the sample. Therefore, what researchers do with their data is what distinguishes between descriptive and inferential statistics (Shende, 2012).

The table given below lists the differences between descriptive statistics and inferential statistics.

Descriptive Statistics	Inferential Statistics
Descriptive statistics are used to quantify and summarize the characteristics of the data.	Inferential statistics are used to make conclusions about the population by applying statistical tests on the sample data.
It is used to describe the characteristics of a known sample or population.	It is used to make inferences about an unknown population
Measures of central tendency and measures of dispersion are the important tests used	Hypothesis testing and regression analysis are the statistical tests used.
Measures of descriptive statistics are variance, range, mean, median, etc.	Measures of inferential statistics are t-test, z test, ANOVA, linear regression, etc.

Descriptive Statistical Tests

Descriptive statistics are used to summarize a set of observations in order to convey the maximum amount of information as simply as possible. This is achieved by applying descriptive statistical tests that describe the basic characteristics of the data in the study. There are three main types of descriptive statistical tests. A distribution represents the frequency of each value. Central tendency represents the mean of the values. Variability or spread describes how the values are spread out (Sarah, 2013).

Distribution Frequency Analysis

Distribution frequency analysis is an important area of statistics that deals with the number of occurrences (frequency) and percentage. It simply counts the number of times each variable occurs, such as the number of males and females within the sample or population.

Measure of Central Tendency

The central propensity measure, also known as the central location measure, is used to determine the representative value of a data set. They provide the values of the distribution. B. Mean, median, and mode, which represent the entire distribution. Representative values of these distributions are compared when researchers wish to make comparisons between two or more of her groups or populations. Central tendency measures facilitate further statistical analysis, as many statistical tests such as variance, skewness, correlation, t-tests, and ANOVA tests are computed using central tendency measures. For this reason, measures of central tendency are also called linear measures. Below are the main dimensions of central tendency.

- Mean - Mean is the mathematical average value of a set of data, calculated using summation of the observations divided by number of observations. It is the most popular measure and very easy to calculate. It is a unique value for a data set, which makes it useful when comparing between data sets. One disadvantage with mean is that it is affected by extreme values (outliers).
- Median - The median, also called positional average, is defined as the middle most observation of the data points when they are arranged either in increasing or decreasing order. Extreme values (outliers) do not affect the median. Like the mean, the median is unique; that is there is only one median of one data set, which is useful when comparing between data sets.
- Mode - Mode is a value that occurs most frequently in a set of observation; that is the observation which has maximum frequency. In a data set, it is possible to have multiple modes or no mode exists, and this renders it not suitable for comparing between data sets (Kothari, 2012).

Measures of dispersion or variation

A measure of variance indicates how distributed the dataset is. It is a quantitative measure of the spread or variability of values in a sample or population. These include variance, standard deviation, standard error, quartile, interquartile range, percentile, range, and coefficient of variation. These are metrics that give an idea about the homogeneity or heterogeneity of your data. The main countermeasures against scattering are shown below.

- Standard deviation and variance - The standard deviation is a measure of how spread out values are from its mean value. It is called standard deviation because we have taken a standard value (mean) to measure the dispersion.
- Standard error - Standard error is the approximate difference between sample mean and population mean. When we draw the many samples from the same population with the same sample size through random sampling technique, then standard deviation among the sample means is called standard error.
- Quartiles and interquartile range - The quartiles are the three points that divide the data set into four equal groups, each group comprising a quarter of the data, for a set of data values which are arranged in either ascending or descending order.
- Percentile - The percentiles are the 99 points that divide the data set into 100 equal groups, each group comprising a 1% of the data, for a set of data values which are arranged in either ascending or descending order. About 25% percentile is the first quartile, 50% percentile is the second quartile also called median value, while 75% percentile is the third quartile of the data.
- *Coefficient of Variation* - Interpretation of standard deviation without considering the magnitude of mean of the sample or population may be misleading. To overcome this problem, coefficient of variation gives an idea. CV gives the result in terms of ratio of SD with respect to its mean value, which expressed in %. CV = 100 × (SD/mean).
- Range - Difference between largest and smallest observation is called range. If A and B are smallest and largest observations in a data set, then the range (R) is equal to the difference of largest and smallest observation, that is, R = A−B.

4. Inferential Statistical Tests

Inferential statistics

In many cases, it is too difficult or too expensive to collect data from the entire population, so researchers can only obtain data from a sample drawn from the population. Descriptive statistics can only summarize sample characteristics, while inferential statistics use sample characteristics to make reasonable inferences about larger populations. Inferential statistical testing draws conclusions from data, generalizes about populations by examining a sample, and makes predictions about the future (Lapan, 2012). Difference statistical tests describe and summarize the properties of datasets. This forms the basis on which inferential statistical tests are performed.

4.1. Types of Inferential Statistical Tests

Inferential statistics have two main uses: making estimates about populations and testing hypotheses to draw conclusions about populations.

4.2. Hypothesis Testing

Statistical analysis using statistical tests is about testing theories and hypotheses. A statistical test is always a test of the researcher's null hypothesis. In hypothesis testing, researchers interpret or interpret given sample data by evaluating two mutually exclusive statements about the population to determine which statement is best supported by the sample data. trying to conclude. More specifically, it tests the probability that the null hypothesis is valid. There are five main steps in hypothesis testing:

Step 1 – The researcher states the hypothesis as a Null and Alternate hypothesis

Step 2 – Chooses a significance level (also called alpha or α)

Step 3 – Collects data in a way designed to test the hypothesis

Step 4 – Performs an appropriate statistical test: compute the p-value and compare from the test to the significance level

Step 5 - Decides whether to *reject* the null hypothesis or *fail to reject* the null hypothesis

Hypothesis testing can be likened to a court case. The Court have to decide whether a person is guilty or innocent based on evidence presented to them. In Court we have:

The Null hypothesis: The person is innocent

The Alternative hypothesis: The person is not innocent.

The Null can only be rejected if there is enough evidence to disprove it.

Forms of Inferential Statistical Tests

Inferential statistical tests come in three forms: tests of comparison, correlation or regression (Murphy, 2019).

Comparison Tests

Comparison tests assess whether there are differences in attributes between two or more groups. The most commonly used statistical comparison tests founded on the assumption that the data is normally distributed include the Z-test, T-test and ANOVA (Analysis of Variance).

Z Test

A z test is a powerful and valuable test that helps the researcher to test if the means of the sample and population are equal when the population variance is known. Suppose a professor claims that the mean score of students in her class is greater than 82 with a standard deviation of 20. If a sample of 30 students yields a mean score of 90, then a Z-test can performed to check if there is enough evidence to support the professor's claim.

Standard t-test

A standard T-test examines whether there is a statistically significant difference in the mean of two independent groups. This is the most basic type of statistical test for use when the researcher is comparing the means from exactly two groups, such as the *control group* versus the *experimental group*. It is often used for continuous independent variables. For instance, a researcher may conduct an experiment studying the effect of a new herbicide on the growth of tomatoes. There are two groups of tomato plants: an experimental group that has been sprayed with the herbicide, and a control group that has not been sprayed. After the researcher has calculated the average growth for each of the two groups, he or she performs a T-test to decide if there is a *statistically significant* difference in their growth (Leavy, 2017).

Another variation of the T-test is the Repeated or Paired T-tests. This is an extremely powerful and most sensitive test for detecting differences. It is sometimes referred to as the dependent sample T-test, and is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. Like the simple T-test, the Paired T-test is often used for continuous independent variable. Like the name suggests, a Paired samples T-test can only be used when the data is paired or matched. Either there is before and after measurements of the same variable, or the T-test can be used to compare how a group of subjects perform under two different test conditions.

4.2.1.1.Analysis of Variance (ANOVA) Test

Unlike the T-test that compares the mean of two groups, the ANOVA test is used to compare the mean of three or more groups. It can be thought of as an extension of the T-test to 3 or more independent groups. T-tests can only compare two groups at a time, and for statistical reasons it is generally considered *illegal* to use t-tests over and over again on different groups from a single experiment (Murphy, 2019).

A One-way ANOVA uses continuous independent variable and is used when groups to be compared are defined by just one factor. Repeated measure ANOVA is used when groups to be compared are defined by multiple factors. For example, if the researcher wants to evaluate the effect of three different antihypertensive drugs on three different groups of human volunteers, then he or she will use ANOVA test to determine any significant difference between related and not independent groups.

Two-way ANOVA is the only way researchers can compare the means of two or more groups according to two different independent variables. The main purpose of two-way ANOVA is to understand whether there is an interaction between two independent and dependent variables (Murphy, 2019). For example, the researchers used her two-way ANOVA to determine whether there was an interaction between physical activity level and gender for blood cholesterol concentrations in adults with low, moderate, and high physical activity levels. I understand. Male or female sex is the independent variable and cholesterol concentration is the dependent variable.

Chi-square Test

The chi-square test is a nonparametric statistical technique that uses dependent variables to determine the difference between observed and expected data. The observed value is the value collected by the researcher and the expected value is the expected frequency based on the null hypothesis (Murphy, 2019). It is useful for investigating whether the difference between two categorical variables is due to chance or the relationship between them. Suppose a researcher wants to know whether gender is independent of education level, then take a random sample, ask each person to indicate the highest level of education attained, and indicate gender. The researchers then perform a chi-square test to determine whether there is an association between a person's gender and the level of education attained.

There are two caveats researchers should be aware of when using the chi-square test. The test is very sensitive to sample size, so a nonsignificant relationship can appear statistically significant with a sufficiently large sample. Chi-square can only determine whether two variables are related. One variable does not necessarily have a causal relationship with another variable. More detailed analysis is needed to establish causality.

Correlation Tests

Correlation tests look for an association between variables checking whether two variables are related. The Chi-square test of independence is the only test that can be used with qualitative variables.

Pearson’s Correlation Coefficient Test

Pearson's correlation coefficient is a parametric test that measures the statistical relationship or association between two continuous variables. The correlation coefficient test assumes that cases are independent and that the two variables are linearly related (Vanderstoep, 2009). For example, if a researcher wants to know if there is a linear relationship between weight and blood pressure, a correlation test can be performed. Correlation only shows the relationship between two variables, not causation.

The main advantage is that correlation tests allow researchers to analyze relationships between many variables in a single study. Correlation coefficients provide a measure of the degree and direction of the relationship. It can reveal the presence or absence of a relationship between two factors and provide guidance for further investigation and research. A major drawback of correlation studies is that the correlation between two variables may be the result of external sources. Researchers should exercise caution and remember that correlations do not always reveal cause and effect. This is also an advantage. If two variables appear related, this can be used as a starting point for more advanced research and experimentation.

4.2.2 Regression Tests

One of the most common and useful statistical tests, regression testing, is used to test causality. This parametric test is used to find out the dependence relationship between two variables. Regression numerically describes the relationship between variables and allows us to predict the dependent variable (y) given the independent variable (x). Regression produces the best fit line according to the traditional equation for lines. $y = mx + b$. An example of a research situation where linear regression can be applied is determining whether a person's weight is linearly related to height (Sherri, 200). Another example is looking at the relationship between a company's monthly income and its advertising spend. The advantage of linear regression analysis is that it is easy to implement and very easy to understand and interpret. Linear regression fits almost perfectly to linearly separable data sets and is often used to determine the nature of relationships between variables. However, linear regression assumes a linear relationship between input and output variables, so it does not fit well to complex data sets. In most real-world situations, the relationships between variables are not linear, so a straight line does not fit the data well.

Non-parametric Tests

A nonparametric test in statistics is a statistical procedure that does not require a distribution to satisfy three basic assumptions, especially when the data are not normally distributed. Nonparametric tests, often called nonparametric tests, are used as an alternative to parametric tests, not as a replacement for parametric tests. In other words, if the data meet the assumptions necessary to perform a parametric test, the appropriate parametric test should be applied. In some cases, researchers can apply parametric tests instead of nonparametric tests, even if the data do not meet the required assumptions and the sample size of the data is large enough.

In order to get correct results from statistical analysis, we need to know when it is appropriate to apply nonparametric tests. The main reason for using nonparametric tests is, first, that the underlying data do not meet assumptions about the population sample. For example, some datasets may have skewed distributions, making the mean no longer the best measure of central tendency, making parametric tests less meaningful. Second, nonparametric tests are performed when the population sample size is too small. If your sample size is too small, you may not be able to verify the distribution of your data. So using nonparametric tests is the only viable option. Finally, nonparametric tests are performed to analyze ordinal or nominal data. Unlike parametric tests, which work only on continuous data, nonparametric tests can be applied to other types of data, such as ordinal and nominal data. For these types of variables, nonparametric tests are the only suitable solution (Murphy, 2019).

The only nonparametric test researchers are most likely to encounter is the chi-square test. However, there are a few others. B. Kruskal-Willis test, a nonparametric alternative to one-way ANOVA. The Mann-Whitney test is a nonparametric alternative to the two-sample t-test. Friedman's test is used to test differences between groups with an ordinal dependent variable. Goodman Kruska's gamma is an association test for ranking variables. On the other hand, the Mann-Kendall trend test looks for trends in time series data. Spearman rank correlation is used when a correlation is found between two data sets.

CONCLUSION

Classifying different types of data variables and presenting them in tables or graphs is an essential and important step in any research. Acquiring this skill will help you integrate research results and prevent the misuse and abuse of tables and figures in research papers. Data, once categorized and summarized, is not an end in itself. Statistical testing must be undertaken to facilitate research conclusions and recommendations. Therefore, data summarization and presentation and data analysis with statistical tests are two sides of the same study of data analysis. One cannot be used without the other.

ACKNOWLEDGMENT

Greatly appreciated is the assistance provided by the DBA132 Research Methodology II Course Supervisor and the Dean in the Department of Business Administration, School of Business, the Chief Librarian of ZCAS University, my fellows students, my family and friends during the course of this study.

REFERENCES

1) Duquia RP (2014), Presenting data in tables and charts. An Bras Dermatol
2) Flockton, L., Crooks, T, and Gilmore, A. (2004). Graphs, Tables and Maps: Assessment Results 2003. Dunedin: Educational Research Unit, University of Otago.
3) Gott, R. and Duggan, S. (2003). Understanding and Using Scientific Evidence: How to Critically Evaluate Data. London: SAGE Publications.
4) Kothari C.R. (2004), Research Methodology - Methods and Techniques, New Age International (P) Ltd, New Delhi, India
5) Lapan S. (2012), Qualitative research - An Introduction to Methods and Designs; John Wiley & Sons, Inc. San Francisco, CA
6) LCMS, (2015); 2015 Living Conditions Monitoring Survey Key Findings; Zambia Central Statistics Office
7) Leavy P. (2017), Research Design; Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches; The Guilford Press, New York, NY
8) Murphy S. (2019), Probability and Statistics; A John Wiley & Sons, Ltd., Publication
9) Nkandela, A. (2001) Rapid Incidences of Child Abuse in Zambia, Lusaka: CHIN.
10) Sarah J. T. (2013), Qualitative research methods - Collecting Evidence, Crafting Analysis, Communicating Impact; A John Wiley & Sons, Ltd., Publication
11) Scribbr (2020), Choosing The Right Statistical Test - Types And Examples. [online] Available at: https://www.scribbr.com/statistics/statistical-tests
12) Shende T (2010), Statistical Methods in Research: Chand and Companny, New Delhi, India
13) Sherri L. J (2009), Research Methods and Statistics - A Critical Thinking Approach, Wadsworth, Cengage Learning
14) Statistics How To. 2020. *Non-Normal Distribution — Statistics How To*. [online] Available at: https://www.statisticshowto.com/probability-and-statistics/non-normal-distributions
15) Upagade V. (2012), Research Methodology; Chand and Company, New Delhi, India
16) Vanderstoep, S. (2009), research methods for everyday life , Blending Qualitative and Quantitative Approaches - Published by Jossey-Bass A Wiley Imprint 989 Market Street, San Francisco, CA 94103