Theoretical Review: An Explanation of Data Types, Statistical Tests, and Factors That Influence Presentation of Findings

Blessed Mwansa, MBA[1] Austin Mwange, DBA[2] Windu Matoka, DBA[3] Joseph Chiseyeng'I, Ph.D.[4]
Onward Chibawe, MBA[4] Robbson Manda, MPA[4] Golden Mashiri, MSc.[4] Richie Nawa, MSc[4]
Nizah Mutambo, MSc[4]
1 Doctorate in Business Administration Student, Department of Business Administration, School of Business, ZCAS University,Lusaka, Zambia
2 Senior Lecturer, Department of Economics, School of Social Sciences, ZCAS University, Lusaka, Zambia, Email: lecturer.researcher@gmail.com
3Senior Lecturer, Department of Business Administration, School of Social Sciences, ZCAS University, Lusaka, Zambia
4 Doctorate in Business Administration Student, Department of Business Administration, School of Business, ZCAS University, Lusaka, Zambia

Abstract
This paper discussed the key factors to consider when selecting the most appropriate tables, and diagrams in the presentation and analysis in research with the view of ensuring data represents what it is intended to represent. Further, the outlines the determinants of the selection of the most appropriate statistics to describe individual variables and to examine relationships between variables and trends in research data. This was desktop research that relied on the use of common databases that scholars use to access free online articles, and these included: Google Scholar, ResearchGate, Academia, and the main Google search engine. The journal articles on research methods, particularly, quantitative research methods were extensively downloaded and used in this study. The key statistical tests discussed in this study include Correlation and Regression. The types of correlation addressed in this paper are: Pearson's Correlation Coefficient; Ranked Correlation Coefficients (Spearman's Rank Correlation Coefficient, and Kendall's Tau Rank Correlation Coefficient); Partial Correlation. For Regression, the two common types have been addressed - Linear Regression, and Logistic Regression.
Keywords: Statistical Tests, Correlation, Pearson's Correlation Coefficient; Ranked Correlation Coefficients, Spearman's Rank Correlation Coefficient, Kendall's Tau Rank Correlation Coefficient, Partial Correlation, Regression, Linear Regression, and Logistic Regression
DOI: 10.7176/RHSS/12-24-01
Publication date: December 31st 2022

1.0 INTRODUCTION

A typical paper uses data management and analysis as a logical narrative to convince the audience that the results of the paper are valid and supported by evidence. From this perspective, this article can be seen as a series of four parts. First, determinants in choosing the most appropriate tables and graphs for the study to ensure that the data represent what was intended. Second, exploring different aspects of data (types of data) in research. Third, considerations to consider when choosing an appropriate statistical test to explain individual variables. Fourth, the study of relationships between variables.

1.1 Objectives: The objectives of this paper were to discuss:

1. determinants of the selection of the most appropriate tables, and diagrams in research to ensure that data represent what it is intended to represent;
2. different aspects of data in research; and
3. factors in the selection of the most appropriate statistical tests to describe individual variables and to examine the relationships between variables and trends in one's data for research.

2.0. Determinants of the selection of the most appropriate tables, and diagrams in Research.

In its raw state, i.e. before H. processing and analysis, it is of little significance (Saunders, 2016). Data analysis is therefore the process of turning data into a narrative and analyzing it to derive meaning (QuestionPro, 2022). It is a process in which a large amount of accumulated data, derived from the above, acquires order, structure, and meaning.

Almost all company research can be categorized as quantitative, qualitative, or mixed methods and either contain some numerical data or data that can be used quantitatively and qualitatively to answer research questions and achieve research goals. (Creswell, 2014). It is important for researchers to have a working knowledge of when to use tables, charts, and statistical approaches if they want their analysis to be useful. Correlations and patterns within specific dates can be examined, plotted, explained and explored using analytical

Research on Humanities and Social Sciences www.iiste.org
ISSN 2224-5766 (Paper) ISSN 2225-0484 (Online)
Vol.12, No.24, 2022

tools such as tables, graphs and statistics (McCullagh, 2002).

It is also important to remember that the purpose of creating easy-to-understand tables, charts and graphs is to facilitate the analysis and interpretation of data, presenting complex problems and situations quickly and effectively.

2.1.1 Quantitative Techniques

Quantitative data can be represented in graphs and charts in a variety of statistical formats and can be easily manipulated statistically. In data analysis, the term "graph" has a unique meaning. However, Saunders (2016) points out that the term is often used interchangeably with "chart". Sanders further described a graph as a visual representation that emphasizes the relationship or relationships between numbers or data elements.

For quantitative analysis, Creswell (2014), tables or graphs show the frequency of occurrence, and the investigator's decisions are influenced by the data element on which the study is focused and the scale of measurement used to collect the data. increase. Most analysis software packages include instructions for creating tables and graphs.

2.1.2 Key Determinants of Utility of Graphs/Charts and Tables

According to Annan (2019), the type of data contained in a dataset has a large impact on researchers' decisions on whether to use graphs/charts or tables. The following topics and subtopics explore key determinants based on different types of data:

2.1.2.1 Continuous Data

a) Histogram

Histograms are used in data analysis to show the highest and lowest values of continuous data. The area of each bar indicates the frequency of occurrence, and the lack of gaps between bars emphasizes the continuity of the data. Data is often divided into class intervals.

b) Frequency Polygon

Is frequently employed when a researcher wants to draw attention to trends and data recall. Quantitative data is represented using a line graph.

c) Pictogram

If the researcher chooses to analyze the data in image format, in the i.H. graph the data is represented by an image or series of photographs instead of each bar. Thumbnail images can be displayed horizontally or vertically. The height of the photos in the columns, or the length of the bar chart, indicates how common the photos are.

While researchers can easily and accurately convert bar charts and histograms to pictograms using today's data analysis tools, it is more difficult to determine the actual data values from the pictograms. This is because it is not always clear how many units a particular part of the image represents. Another drawback is that pictograms can easily be tampered with. When using icons, researchers should choose a standard value for each image and maintain consistent image sizes to compensate for this error. (Anna, 2019).

d) Line Graphs

It can be applied to variables with numerical and temporal data to show trends. Continuous and discrete time trends can also be displayed using histograms and bar charts respectively.

e) Pie Charts

Pie charts are the most commonly used graphics for highlighting percentages of events. It is divided into segments proportionally based on the percentage of total value each section contributes and the total value represented by the circle is recorded. A cluttered pie chart (more than 6 slices) is difficult to understand, so researchers need to sort numerical data and some categorical data before creating a pie chart.

f) Bar Charts

Bar charts can be divided into several main categories, such as stacked, component, compound, and many bar charts. They can be used to compare ratios, trends, cumulative totals, etc.

Key traits are summarized in the table below:

Trait	Categorical		Numerical		References (secondary)
	Descriptive	**Ranked**	**Continuous**	**Discrete**	
To show one variable so that any specific amount can be read easily	Table/frequency distribution (data is often grouped)				Evans, Gruba, & Zobel, 2014; Creswell, 2009

To show the relative amount for categories or values for one variable so that highest and lowest are clear	Bar graph/chart or pictogram (data may need grouping)		Histogram or frequency polygon (data must be grouped)	Bar graph/chart or pictogram (data may need grouping)	Creswell, 2014; Blumberg, Cooper, & Schindler, 2014; Annan, 2019
To show the trend for a variable		Line graph or bar graph/ chart	Line graph or histogram	Line graph or bar graph/chart	Blumberg, Cooper, & Schindler, 2014
To show the proportion or percentage of occurrences of categories or values for one variable	Pie chart or bar graph/chart (Data may need grouping)		Histogram or pie chart (data must be grouped)	Pie chart or bar graph/ chart (data may need grouping)	Evans, Gruba, & Zobel, 2014; Creswell, 2009; Annan, 2019
To show the distribution of values for one variable			Frequency polygon, histogram (data must be grouped) or box plot	Frequency polygon, bar graph/chart (data may need grouping) or box plot	Creswell, 2014; Creswell, 2009; Annan, 2019
To show the interdependence between two or more variables so that any specific amount can be read easily	Contingency table/cross-tabulation (data often grouped)				Creswell, 2014; Blumberg, Cooper, & Schindler, 2014
Comparing the relative amount for categories or values for two or more variables so that highest and lowest are clear	Multiple bar graph/chart (continuous data must be grouped, other data may need grouping)				Evans, Gruba, & Zobel, 2014; Creswell, 2009
To compare the proportions or percentages of occurrences of categories or values for two or more variables	Comparative pie charts or percentage component bar graph/chart (continuous data must be grouped, other data may need grouping)				Creswell, 2014; Creswell, 2009
To compare the distribution of values for two or more variables			Multiple box plot		Creswell, 2014; Dancey & Reidy, 2011
To compare the trends for two or More variables so that intersections are clear		Multiple line graph or multiple bar graph/chart			Evans, Gruba, & Zobel, 2014); Creswell, 2014
Comparing the frequency of occurrences of categories or values for two or more variables so that cumulative totals are clear	Stacked bar graph/chart (continuous data must be grouped, other data may need grouping)				Creswell, 2014; Blumberg, Cooper, & Schindler, 2014; Dancey & Reidy, 2011

To compare the proportions and cumulative totals of occurrences of categories or values for two or more variables	Comparative proportional pie charts (continuous data must be grouped, other data may need grouping)	Evans, Gruba, & Zobel, 2014; Blumberg, Cooper, & Schindler, 2014
To show the interdependence between cases for two variables	Scatter graph/scatter plot	Creswell, 2014; Dancey & Reidy, 2011

key themes adapted from Saunders, Lewis, and Thornhill (2015)

2.2. An Exploration of Different Types of Data in Research

Any type of data has the extraordinary built-in ability to describe an object after it has been assigned a particular value. To make these values relevant to their research, researchers need to organize them in a specific context. Most data can be divided into qualitative and quantitative categories (QuestionPro, 2022). Many business statistics publications use measurement hierarchies to classify quantitative data into different data types, often in increasing order of numerical precision (Saunders, Lewis, & Thornhill, 2016).

When analyzing quantitative data, it is important to be aware of the differences between different forms of data. This is because it is easy to obtain statistics from data sets that are incorrect for the relevant data type, providing little or no useful information (Evans, Gruba & Zobel, 2014). Annan (2019) further points out that the more precise the measurement scale, the greater the range of analytical techniques available to researchers. Evans (2014) states that understanding different kinds of data is essential for information. This includes understanding how the data was collected, where it came from, what characteristics were measured, and what the data consists of. Similarly, the type and source of data influence the determinants described in (3.1) above. Simply put, factors influence the behavior of data so that researchers can accurately predict how variables and data values will interact in a study.

2.2.1 Qualitative Data Types

Evans (2014) observed that, qualitative data can be classified using the measurement of scale as nominal (classificatory in nature) and ordinal (ranking or scale in nature). These are further explained below:

(i) Categorical data

Refers to data whose values can be classified or grouped into groups (categories) based on characteristics that identify or describe a variable, but that cannot be numerically quantified. They can also be classified into descriptive and ordered categories (Marshall & Rossman, 2011). Because categories cannot be numerically quantified or ranked, they are called descriptive or nominal data. Instead, these statistics count only instances of the variable in each category. Categories should be unique and unambiguous in most analyses, or have only one attribute.

In addition, if the variable is split into two categories, for example, if the variable's gender is split between male and female (there are only two categories), we distinguish descriptive data as dichotomous. Additionally, ordinal data is a more precise type of categorical data that includes rating or scale questions (Blumberg, Cooper, & Schindler, 2014).

Nominal data can be primary and secondary data. This includes symbols, letters, words, gender and other types of information. Nominal data are evaluated according to the grouping approach. This method categorizes the data, calculates the frequency or percentage of the data, and graphs it using a pie chart.

Ordinal data, on the other hand, shows variables and data that follow their natural order. This variable is common in surveys, finance, marketing, and other fields. Ordinal data is often displayed using bar charts. It can also be viewed as a table with each row representing a separate category.

(ii) Quantitative Data Types

Numeric (discrete and continuous) and interval scales can be used to categorize quantitative data into two separate groups. These are covered in more detail below:

(a) Numerical data:

These are items whose values can be quantified or counted (Berman Brown and Saunders 2008). Numeric data is more precise than categorical data because each data value can be associated with a position on the numeric scale. Furthermore, the variable nature means that a wider range of statistics can be used to examine this data.

(b) Continuous and Discrete Data

According to Dancey and Reidy (2011), continuous data are features whose values can theoretically take on any value. In addition, we have individual data, where each case selects a value from a scale that measures change in individual units.

According to Blumberg (2014), interval data represent the difference or "interval" between any two data values of a given variable, if there is no guarantee of relative difference. neither division nor addition or subtraction of them in any meaningful way is possible.

2.3 Factors to Consider in the Selection of the Appropriate Statistical Tests to Describe Individual Variables.

2.3.1 Descriptive statistics

This statistical method, according to Creswell (2014), demonstrates the essential traits of various types of data used in research. Descriptive statistics help to meaningfully present the data in a pattern that makes sense. The following are some of the major types of descriptive analysis methods:

(i) Measures of Central Tendency

These include the mean, mode and median. And are further discussed below:

(a) Mean

The term "average" refers to a single value that is considered the most common or representative value for a given data set. This is the value at which the data in the set tends to be concentrated. It is also calculated by summing all values in a sample or population and multiplying by the total number of values. Before using averaging, it is important to ensure that the data are symmetrical, or at least not significantly skewed. This is because for symmetric data the mean provides the closest approximation to the center of the distribution. On the other hand, if your data are highly skewed, do not use the mean to determine the center of the distribution (Dancey & Reidy, 2011).

(b) Mode

Creswell (2014) argues that modes are the most frequently occurring observations. The most common observations in a dataset are where to look for modes in that dataset. This mode is most useful for determining the most frequently observed observations and when quick estimation of central tendency is required. In contrast, Dancey (2014) states that this mode should not be used if the data are multimodal, highly skewed, or uniform. This is because the fulfillment center estimate can be very inaccurate under these circumstances. Also, it should not be used when mean, median, or other more accurate measures of central tendency are available.

(c) Median

The median of the data is the middle observation. That is, half of the data are below the median and half are above the median. To determine the median, the data must be ordered from smallest to largest observation. The median is a great tool when researchers need to determine whether new data points are above or below the midpoint, when the data are highly skewed, or when there are outliers that change the mean. is.

Basically, the choice of which approach to use should depend on the appropriate use of the central tendency measure and data type. Specifically, the mean for interval level/ratio variables, the median for ordinal variables, and the mode for nominal variables. The "closeness" of the mean, median, and mode is further influenced by the shape of the distribution and the presence of outliers.

The mean, median, and mode will typically have values that are near to one another when a distribution is symmetric and free of outliers.

When a distribution is skewed to the right, the relationship between mean, median and mode is usually described by:

Mode < Median < Mean (mode is the smallest and mean is the largest).

When a distribution is skewed to the left, the opposite is generally true:

Mean < Median < Mode (mean is the smallest and mode is the largest).

Basically, descriptive analysis is often used in quantitative market research to provide absolute statistics, but rarely is the analysis sufficient to justify these figures. However, it is important to consider the research and data analysis techniques that are appropriate for a particular study, as well as the story researchers want to tell.

It is preferable to rely on descriptive statistics when researchers attempt to keep studies and results specific to a particular sample without generalizing. Because descriptive analysis is often used to assess a single variable, it is also called "univariate analysis".

2.4 Examining Relationships, Differences, and Trends Using Statistics

This section describes statistical techniques that can be used to examine relationships between two or more variables. The first part considers the relationship between two variables (bivariate association). Additionally, it highlights a selection of statistical techniques for assessing variable relationships, shows how relationship analysis can be used to make predictions, and provides advanced statistical relationship analysis used in research.

2.4.1 Examining Relationships Using Statistics

When the terms "relationship" and "association" are used, they mean that two variables change or fluctuate together. Two variables can change at the same time, but the change is not necessarily statistically significant. So the two types are significant and non-significant variation or correlation.

A correlation between two variables that is statistically significant is called a statistical relationship. This significance is based on the level of the probability test, or the p-value of the Pearson's correlation coefficient. When one variable increases or decreases, another variable follows. If the variability is due to something other than chance, it is statistically significant (Rosner, 2016). According to Devonish (n.d.), common inferential statistics used to analyze associations between variables include correlation, regression (single and multiplicative linear), and chi-square tests. These are detailed below:

(a) Correlations

This method is commonly used when researchers want to study relationships between two or more variables. A correlation is a statistical relationship between two variables. Pearson's correlation (Devonish, undated) can be used to determine the strength and direction of the relationship between two numerical or quantitative variables. It should be emphasized that correlations can only be used to describe quantitative variables. This is due to the lack of means and standard deviations for categorical variables.

Researchers compute a correlation coefficient and a coefficient of determination to ascertain the statistical association between two variables.

(i) Correlation Coefficient:

The strength and direction of the linear relationship is measured by the correlation coefficient. It is calculated using the mean and standard deviation of the x and y variables. We turn our attention to the nature of the relationships between variables and the numerical breakdown of their ranges. The formula for the correlation coefficient is r = +/-x. where r is the correlation coefficient, a and b are the two correlated variables, and the plus or minus sign indicates the direction (positive or negative) of the relationship between the variables. , where x is a specific value.

The sign (+ or -), which denotes whether the relationship between the variables of interest is positive or negative, is the initial component.

A correlation value of 0.00 denotes that there is no relationship between two variables, at least linearly. The second component is a numerical value that expresses how strongly the variables are related to one another. A decimal value is used to represent this number, and it runs from +1.00 (the ideal positive connection) to -1.00. (a perfect negative relationship).

(ii) Interpreting correlation coefficients: The significance or strength of a correlation coefficient depends on several factors, including the goal and application of the study as well as the size of the sample.

(iii) Calculating correlation coefficients: Researchers utilize a variety of statistical approaches to calculate correlation coefficients between two variables, depending on how the two variables are measured

The relationships between ratio/interval variables can be investigated using the following methods.

(a) The Pearson product moment correlation creates a correlation coefficient for two variables that are assessed on a ratio or interval scale.

(b) When researchers assess one variable using a ratio/interval scale and the second variable using a nominal scale, they employ the point biserial correlation (rpb).

It is worth noting that the relationships between ordinal variables can be assessed as follows:

a) **The Spearman correlation:** using the Spearman correlation, a researcher can compare two sets of ranked scores for the same group of study traits or compare the ranked scores of individual items by two different groups.
b) **Kendall's correlation:** is often employed when a researcher has two ranks for each of several different study traits.

(iii) Correlation Matrices: Researchers use asterisks (eg, asterisks indicate significance at the 0.05 level) or notes included in the matrix to indicate bivariate correlation coefficients and their significance. The correlation matrix lists all variables associated with the left-hand top and bottom of the matrix where each row and column intersect. Me. e. Correlation and Causation Although causality cannot be automatically inferred from correlation coefficients, one of the factors used to determine causality is correlation.

When attempting to infer causation, researchers occasionally consider the chronological order of occurrences.

Two variables may also be associated, but the significance of this relationship is not always clear.

(iv) Coefficient of Determination: The coefficient of determination is a mathematical expression of how much variance in one variable is related to, explained by, or determined by another (r-squared). Squaring the correlation coefficient, which ranges from 0.00 to 1.00, yields the coefficient of determination. Researchers must pay particular attention to the correlation of determination while assessing the correlation coefficients found in their investigations.

The coefficient of determination is a formula for how much the variance of one variable is related to, explained by, or determined by another variable (r-squared). Squaring the correlation coefficient between 0.00 and 1.00 gives the coefficient of determination. Researchers should pay particular attention to decision correlations when evaluating correlation coefficients found in studies.

(v) Multiple Correlation:
When assessing the association between the variable they are trying to explain, the reference variable, and two or more additional independent variables that work together, researchers calculate multiple correlations that produce two different types of statistics.

(a) The multiple correlation coefficient ® is the same as the correlation coefficient, except that it informs the researcher about the relationship between two or more variables and the target criterion variable. Shows the relationship between a reference variable and other variables in terms of both strength and direction. Read "Multiple correlations between variables b and c" in the form Ra.b.x = +/-x.
(b) The amount of variance in a reference variable that can be explained by interactions with other variables is represented by the multiple coefficient of determination (R-squared, R2), which is calculated by the square root of the multiple correlation coefficient. The portion of variation in the reference variable that the independent variables cannot explain is called the non-determined coefficient and is denoted by the symbol 1-R2.

(v) Partial Correlation: Partial Correlation: A partial correlation statistically accounts for the impact of one or more additional variables while explaining the link between two variables (sometimes called effects analysis or elaboration).
Essentially, it is crucial for a researcher to comprehend that correlations have a variety of strengths, which are highlighted as follows:

(a) 10 to .29 = Weak correlation/relationship
(b) .30 to .49 = Moderate relationship/Medium correlation
(c) .50 and above = Strong relationship/high correlation.

The sign of the relationship does not indicate the strength; for example, (-).50 is the same strength as (+).50 but different direction. ‘r’ is the symbol of the correlation coefficient.

Regression Analysis
Regression analysis is used by researchers to understand significant associations between two or more variables. Identification of dependent and independent variables is essential in this methodology (Creswell, 2014). When a study attempts to determine how independent variables affect dependent variables, it assumes that the values of the independent and dependent variables are completely randomly generated (Creswell, 2009).

Both numeric variables and dichotomous variables (categorical variables with only two categories or groups) should be used as independent or predictor variables. On the other hand, the dependent variable must be numeric or quantitative (Devonish, nd). Simple regression and multiple regression are the two major subcategories of regression. Simple linear regression explores the relationship between predictor and outcome variables, while multiple regression explores the effects of multiple predictor or independent variables on a single outcome variable. It is important to understand that both have the same meaning and lead to the same result.

Regression analysis is performed by creating regression equations, sometimes called regression models, which are algebraic equations that describe the relationships between variables. A typical two-variable regression equation is Y = a + bc. where y is the reference variable and intercept, b is the slope, and x is the predictor variable. The slope and intercept each indicate how closely the regression line intersects the y-axis and depend on the correlation between the two variables. The slope indicates how many units the variable Y increases for each unit increase in X. (Frey, Peony, Krebs, 1999).

A significance test is used to assess whether the expected variance of the regression equation is significant. The term "goodness of fit" also refers to how well a model or equation (such as a regression line) describes or "fits" the data.

Multiple Linear Regression
As emphasized above, researchers can use information from two or more predictor variables and an understanding of the relationships between all variables to predict or explain results for the reference variables. Regarding how predictors are introduced into the regression analysis and how they determine how much variance they explain in the reference variables, the following examples further illustrate different ways to perform multiple regression analysis:

Hierarchical regression analysis: The researcher chooses the order in which the variables are entered into the regression equation based on prior theory and study.

Stepwise regression: Using computational techniques, predictor variables are entered into the computer in various arrangements and sequences until the "best" equation is found.

It's vital to remember that multiple linear regression gives researchers access to at least three critical pieces of knowledge, including:

(i) a multiple correlation coefficient® that demonstrates the relationship between each predictor variable and the criteria variable.
(ii) a measure of how much of the variance in the criterion variable can be accounted for by the interactions between the predictor variables known as the coefficient of multiple determination (R2). The number of

independent variables examined is taken into account when calculating the modified R2 (*R2).

(iii) We may determine how much each predictor variable contributes to explaining the criterion variable by providing a regression coefficient, or beta coefficient, that displays the amount, or relative weight, that each predictor variable offers to explaining the scores on the criterion variable.

Advanced Relationship Analysis

Multivariate analytical techniques that analyze relationships between three or more variables are more sophisticated. Here are a few of them in brief:

Canonical Correlation Analysis (Rc): is a type of regression analysis that is used to look at how various independent and dependent variables relate to one another.

Path analysis: explores proposed relationships between several variables (often independent, mediating, and dependent) in order to demonstrate the "paths" that causal influences take in order to create causal connections and inferences.

Discriminant analysis: is a type of regression analysis that assigns different categories to study features based on how well they perform on two or more ratio/interval independent variables.

Factor analysis: determines whether a large number of variables may be broken down into fewer factors (a set of variables).

Cluster analysis: Determines whether several variables or elements may be grouped together into meaningful clusters.

Multidimensional scaling (MDS) displays the statistical similarities and contrasts between and among variables or items in two or more dimensions.

Regression analysis basically examines how different (predictor/independent) factors affect a single (dependent) outcome variable. An important concept in regression analysis is the use of the word "prediction". This is because we need to determine whether one variable predicts, explains, or influences another. Furthermore, we can conclude that the correlations between variables can quickly become numerous and complex, especially if the goal is to predict about something. These relationships range from positive to neutral to negative. Care must therefore be taken in determining the difference between correlation and causation. Understanding is also required to interpret statistical correlations between variables.

Chi-Square Test

Pearson Chi-Square Test

Pearson's chi-square test examines whether there is a statistically significant association between two categorical variables (Gordon, 2004). It is used in conjunction with a crosstabulation of two variables. Researchers should be aware of the categorical independent and dependent variables required to perform this function.

The following significant questions must be addressed before performing a Chi-square test:

(a) How many variables are there?
(b) What distinguishes independent variables from dependent variables?
(c) Which variables fall under which categories?

2.4.2 Examining Differences Using Statistics

We often want to compare groups of samples. When comparing sets of samples, researchers must consider measurement variability. Below are some of the many statistics introduced by this logic that can be used to assess sample differences:

Hypothesis testing: The probability that a given hypothesis is true is determined by hypothesis testing (Gordon, 2004). This method of analysis shows the relationships between multiple variables rather than focusing on a single variable. To analyze relationships between variables, researchers typically use it when they need something other than absolute numbers (Crawley, 2012). Hypothesis testing uses information from a sample to compare a null hypothesis to an alternative hypothesis for the value of a parameter within a population. The alternative hypothesis asserts that the null hypothesis is false, false, or false in a certain way, while the null hypothesis often makes certain statements about the parameters (DeCoster, 2006).

According to DeCoster (2006), the determination of hypothesis testing often follows this pattern:

(a) Identify the null hypothesis and the alternative hypothesis.
(b) Select a sample from the target demographic.
(c) Gather information that will allow you to distinguish between the null and alternative hypotheses.
(d) Determine a test statistic using the information.
(e) Calculate the p-value of your test statistic, which is the likelihood that you would have obtained an extreme test statistic if the null hypothesis were true.

It is important to note that the distribution from which the test statistic was derived must be known to the researcher in order to obtain the p-value. Note that the overall form of the distribution and the degrees of freedom associated with the test statistic are usually used and referred to.

Decision Point:

(a) reject the null hypothesis in favor of the alternative hypothesis if the p-value is low.

(b) If the probability is high, you fail to reject the null hypothesis.

It is important to remember that the significance level is the threshold by which researchers decide to accept or reject the null hypothesis. With a standard deviation of 0.05, there is a 5% chance that the result does not agree with the null hypothesis by sheer chance.

It is also recommended to use exact p-values (p=0.03) instead of confidence intervals (p=0.05) when reporting p-values. Also, if the p-value is greater than or equal to 0.10, we recommend using two valid numbers. Additionally, if the p-value is less than 0.10, you should be careful with significant digits. Finally, the p-value cannot be expressed as equal to 0.000. Instead, it should always be defined as p < .001

Analysis of Variance (ANOVA)

ANOVA is a method of testing whether the means of three or more populations are equal. Among the various types of his ANOVA, the one-way ANOVA is of particular interest in this section. An ANOVA test is performed to determine if there are statistically significant differences between the means of different groups. This test uses variation to determine whether the means are equal (Boyd & Casper, 2021).

Boyd (2021) lists five fundamental presumptions that must be true to execute a one-way ANOVA, including:

(a) It is assumed that every population from which a sample is drawn is normal.
(b) Each sample is independently chosen at random.
(c) It is assumed that the populations' standard deviations are equal (or variances)
(d) The variable is a categorical factor.
(e) The outcome is a number-based variable.

The null hypothesis in this situation is simply that the group population means are all equal, and the alternative hypothesis is that at least one pair of means differ. This should be taken into account when interpreting the results. Therefore, if;

H_0 is true: All means are equal, and the variations are caused by random variation.

H_0 is not true: all means are not the same, and the variations are too great to be the result of random variation.

Ultimately, the approach is applicable if:

(a) All populations of interest have normal distributions.
(b) Standard deviations are the same across populations.
(c) Samples are drawn at random and independently from each population, however they need not all be the same size.
(d) The F ratio is the test statistic for an analysis of variance.

2.4.2 Examining Trends Using Statistics

Trend analysis, which describes patterns over time, is commonly used in various academic disciplines such as business, finance, and economics. The number of recent or historical events and their variability or uncertainty can be estimated using trend analysis. It also serves as a basis for forecasting and predicting, after examining the importance of time and its relationship with other predictors (Chao, Wu, Wu, & Wei-Chih, 2018).

According to Bautista (2008), trend analysis is a statistical technique used to assess both linear and nonlinear correlations between two quantitative variables. It is usually applied as a regression analysis or analysis of variance (ANOVA) for numeric variables. Trend analysis can be used to modify a single independent variable or factor and see how it affects the dependent or response variable, and is generally applicable to data obtained over time (Chao et al., 2018; Bautista, 2008). To statistically determine the form, structure, or course of such relationships, the dependent variable is tracked through states, levels, or points of the manipulated independent variable.

(1) Descriptive Statistics

Trend analysis can be performed using descriptive statistics. This is because descriptive information types are typically used to summarize specific data sets or other statistics collected from larger groups. Measures of variance, such as standard deviation, range, variance, or maximum random variable, as well as measures of central tendency, such as mean, median, and mode, best represent the statistical techniques available (Vitez, 2022). Trend analysis can be used for certain types of information, such as: B. Income, Income, Expenses, and Comparable Financial Data Analysis.

(2) Inferential Statistics

Trend analysis statistics can be performed using inference statistics, which often rely heavily on probability statistics. This style often selects samples from larger populations and, as mentioned above, draws conclusions from huge data sets. Researchers commonly use these statistics to estimate the likelihood that a larger population behaves like the sample.

In addition to linear regression, trend analysis can also look for nonlinear trends using more complex statistical techniques such as the Mann-Kendall test. Similarly, MANCOVA (Multivariate Analysis of Covariance) can be used to determine whether group differences are more random or whether repetitive patterns emerge (Chandler & Scott, 2011).

Although there are many potential uses for trend analysis, there are some potential drawbacks that

researchers should consider before conducting trend analysis. For example, sampling errors affect all data (except when collected as part of a census). Using coarse-grained sampling techniques (such as simple sampling) exacerbates the severity of this problem. Moreover, extrinsic, systematic or random measurement errors are undoubtedly a threat to your data. It is possible to confuse this error trend with the actual data trend.
The summary table below shows the criteria that can be used when choosing a statistical test to use.

Criteria for choosing Statistical test

Trait	N# of independent variables	N# of Dependent Variables	Covariates	Type of independent or dependent variables	Distribution of Score	Statistical test
Group comparison	1	1	0	Categorical/continuous	Normal	t-test
Group comparison	≥ 1	1	0	Categorical/continuous	Normal	Analysis of Variance
Group comparison	≥ 1	1	1	Categorical/continuous	Normal	Analysis of Covariance
Association between groups	1	1	0	Categorical/continuous	Non-normal	Chi-square
Relate variables	1	1	0	Categorical/continuous	Normal	Pearson product moment correlation
Relate variables	≥ 2	1	0	Categorical/continuous	Normal	Multiple regression

adapted from (Cresswell, 2014)

3.0 CONCLUSION

This article aimed to highlight factors that should be considered while selecting the most appropriate tables and diagrams for research, exploring various data characteristics, and choosing the right statistical tests to characterize individual variables and look at links between variables.

The purpose of this article is to help you select the most suitable tables and graphs for your investigation, explore various data characteristics, and find appropriate statistical tests to characterize individual variables and explore relationships between them. It is to highlight the factors to consider when choosing.

In order for researchers to be useful when it comes to examining charts and graphs, it is important to have a working knowledge of when to use tables, charts, and statistical techniques for analysis. It is also important to remember that the purpose of creating easy-to-understand tables, charts, and graphs is to facilitate the analysis and interpretation of data, effectively and quickly communicating complex issues and situations.

Moreover, the ability to correctly apply statistical measures to data and draw certain conclusions from it makes data types an important concept in statistics. Exploratory data analysis (EDA) requires a solid understanding of different data types, as it may only be possible to use certain factual measures of the data type. To choose the best recognition method, you need to know your data analysis and what kind of data analysis you are using. A data type can be viewed as a means of organizing different kinds of variables. The ultimate goal of many statistical techniques used in data analysis in research is to provide accurate and reliable information. To reduce statistical errors, researchers must learn how to deal with common problems such as outliers, missing data, data updates, data mining, and graphing.

Bibliography

Bautista, R. (2008). Trend analysis. In P. J. Lavrakas (Ed.), Encyclopedia of survey research methods . *Sage Publications Inc*, 907-907.

Boyd , R., & Casper, N. (2021). *Introductory Statistics 4th Edition (revised).* California: College of Lake County Faculty.

Chao, Y.-S., Wu, C.-J., Wu, H.-C., & Wei-Chih, C. (2018). Trend analysis for national surveys: Application to all variables from the Canadian Health Measures Survey cycle 1 to 4. *PLos One*.

DeCoster, J. (2006). *Testing Group Differences using T-tests, ANOVA, and Nonparametric.*

Vitez, O. (2022, October 16). *What Are the Different Types of Trend Analysis Statistics?* Retrieved from SmartCapitalMind: https://www.smartcapitalmind.com/what-are-the-different-types-of-trend-analysis-statistics

Annan, D. (2019). *A Simple Guide to Research Writing.* Flash Print.

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2014). *Business Research Methods 4th Edition.* Berkshire: McGraw-Hill Education.

Chandler, R. E., & Scott, M. E. (2011). *Statistical Methods for Trend Detection and Analysis in Environmental Sciences.* Wiley.

Crawley, M. J. (2012). *The R Book.* John Wiley and Sons.

Cresswell, J. W. (2014). *Research Designs: Qualitative, Quantitative and mixed Method Approaches 4th Edition.* London: SAGE Publications.

Creswell, J. W. (2009). *Research Design. Qualitative, Quantitative and Mixed Method Approaches 3rd Edition.* SAGE Publications Inc.

Dancey , C. P., & Reidy, J. (2011). *Statistics Without Maths for Psychology.* Prentice Hall.

Devonish, D. (n.d). Exploring Relationships using SPSS Inferential Statistics:Part II.

Evans, D., Gruba, P., & Zobel, J. (2014). *How to Write a Better Thesis 3rd Edition.* Springer International Publishing Switzerland.

Frey, L., Botan, C., & Kreps, G. (1999). An Introduction to Research Methods. 2nd Edition. Boston: Allyn & Bacon.

Gordon, S. (2004). *Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments." Statistical Applications in Genetics and Molecular Biology.*

Marshall, C., & Rossman, G. (2011). *Designing Qualitative research 5th Edition.* SAGE Publications.

McCullagh, P. (2002). *The Annals of Statistics. Vol 30.* Institute of Mathematical Statistics.

QuestionPro. (2022, October 21). *Data analysis in research: Why data, types of data, data analysis in qualitative and quantitative research*. Retrieved from QuestionPro: https://www.questionpro.com/blog/data-analysis-in-research/

Rosner, B. (2016). *Fundamentals of Biostatistics 8th Edition.* Boston.

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students 7th Edition.* Pearson Professional Limited.