

Review

An Overview of Data Analysis and Interpretations in Research

Dawit Dibekulu Alem

Lecturer at Mekdela Amba University, College Social Sciences and Humanities , Departement of English Language and Literature. Email: dawitdibekulu7@gmail.com

Accepted 16 March 2020

Research is a scientific field which helps to generate new knowledge and solve the existing problem. So, data analysis is the crucial part of research which makes the result of the study more effective. It is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information. In a research it supports the researcher to reach to a conclusion. Therefore, simply stating that data analysis is important for a research will be an understatement rather no research can survive without data analysis. It can be applied in two ways which is qualitatively and quantitative. Both are beneficial because it helps in structuring the findings from different sources of data collection like survey research, again very helpful in breaking a macro problem into micro parts, and acts like a filter when it comes to acquiring meaningful insights out of huge data-set. Furthermore, every researcher has sort out huge pile of data that he/she has collected, before reaching to a conclusion of the research question. Mere data collection is of no use to the researcher. Data analysis proves to be crucial in this process, provides a meaningful base to critical decisions, and helps to create a complete dissertation proposal. So, after analyzing the data the result will provide by qualitative and quantitative method of data results. Quantitative data analysis is mainly use numbers, graphs, charts, equations, statistics (inferential and descriptive). Data that is represented either in a verbal or narrative format is qualitative data which is collected through focus groups, interviews, opened ended questionnaire items, and other less structured situations.

Key Words: Data,data analysis, qualitative and quantitative data analysis

Cite This Article As: Dawit DA (2020). An Overview of Data Analysis and Interpretations in Research. Inter. J. Acad. Res. Educ. Rev. 8(1): 1-27

INTRODUCTION

Research can be considered as an area of investigation to solve a problem within a short period of time or in the coming long future. As explained by Kothari (2004), research in common parlance refers to a search for knowledge. It can also be defined as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English

Oxford, (1952, p. 1069) cited in Kothari (2004), lays down the meaning of research as "a careful investigation or inquiry especially through search for new facts in any branch of knowledge." Moreover, Redman and Mory (1923) cited in Kothari (2004), define research as a "systematized effort to gain new knowledge."

In research getting relevant data and using these data properly is mandatory. The task of data collection begins after a research problem has been defined and research design/ plan chalked out. While deciding about the

method of data collection to be used for the study, the researcher should keep in mind two types of data viz., primary and secondary. The *primary data* are those which are collected afresh and for the first time, and thus happen to be original in character. The *secondary data*, on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection. The methods of collecting primary and secondary data differ since primary data are to be originally collected, while in case of secondary data the nature of data collection work is merely that of compilation. Whatever it is the data used in any research should be analyzed properly either qualitatively or quantitatively based on the nature of the data collected. Data collected from various sources can be gathered, reviewed, and then analyzed to form some sort of finding or conclusion.

There are a variety of specific data analysis method, some of which include data mining, text analytics, business intelligence, and data visualizations. Patton(1990) stated that data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. It can be done qualitatively or quantitatively. Data analysis is the central step in both qualitative and quantitative research.

Whatever the data are, it is their analysis that, in a decisive way, forms the outcomes of the research. The purpose of analyzing data is to obtain usable and useful information. The analysis, irrespective of whether the data is qualitative or quantitative, may: describe and summarize the data, identify relationships between variables, compare variables, identify the difference between variables and forecast outcomes. Sometimes, data collection is limited to recording and documenting naturally occurring phenomena, for example by recording interactions which may be taken as qualitative type. Qualitative analysis is concentrated on analyzing such recordings. On the other hand data may be collected numerically using questionnaires and some rating scales and these data mostly analyzed using quantitative techniques.

With this introduction, this paper focuses on data analysis, concepts, techniques, expected assumptions, advantages and some limitations of selected data analysis techniques. That is, the concept of data analysis and processing steps are treated in the first part of the paper. In second part, concepts of qualitative and

quantitative data analysis methods are explained in detail. More over emphasize is given to descriptive and inferential statistics methods of data analysis. Finally, the how of writing summary, conclusion and recommendations based on the findings gained qualitatively as well as quantitatively are included.

Data Analysis

Concept of Data Analysis

What do we mean when we say data in the first place? The 1973 Webster's New Collegiate Dictionary defines data as "factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation." The 1996 Webster's II New Riverside Dictionary Revised Edition defines data as "information, especially information organized for analysis." Merriam Webster Online Dictionary defines data" as: factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation; information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful or information in numerical form that can be digitally transmitted or processed.

Taking from the above definitions, a practical approach to defining data is that it is numbers, characters, images, or other method of recording, in a form which can be assessed to make a determination or decision about a specific action. Many believe that data on its own has no meaning, only when interpreted does it take on meaning and become information. By closely examining data (data analysis) we can find patterns to perceive information, and then information can be used to enhance knowledge (The Free On-line Dictionary of Computing, 1993-2005 Denis Howe).

Simply, data analysis is changing the collected raw data into meaningful facts and ideas to be understood either qualitatively or quantitatively. It is studying the tabulated material in order to determine inherent facts or meanings. It involves breaking down existing complex factors into simpler parts and putting the parts together in new arrangements for the purpose of interpretation. As to Kothari (2004) data analysis includes comparison of the outcomes of the various treatments upon the several groups and the making of a decision as to the achievement of the goals of research. The analysis, irrespective of whether the data is qualitative or quantitative, may be to describe and summarize the data, identify relationships between variables, compare variables, identify the difference between variables and forecast outcomes as mentioned in the introduction.

According to Ackoff (1961), a plan of analysis can and should be prepared in advance before the actual

collection of material. A preliminary analysis on the skeleton plan show as the investigation proceeds, develop into a complete final analysis enlarged and reworked as and when necessary. This process requires an alert, flexible and open mind. Caution is necessary at every step.

In the process of data analysis, statistical method has contributed a great deal. Simple statistical calculation finds a place in almost any study dealing with large or even small groups of individuals, while complex statistical computations form the basis of many types of research. It may not be out of place, therefore to enumerate some statistical methods of analysis used in educational research. The analysis and interpretation of data represent the application of deductive and inductive logic to the research process.

Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis. The term analysis refers to the computation of certain measures along with

Data Processing Operations

In the data analysis process we need to focus on the following data analysis process operation stages. Kothari (2004) suggested the following data analysis operation stages.

1. Editing: Editing of data is a process of examining the collected raw data (especially in surveys) to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation (Kothari, 2004). So, this indicates that editing is the process of data correction.

According to Kothari (2004) editing should be done, one can talk of field editing and central editing.

Field editing consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and/or in illegible form at the time of recording the respondents' responses. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher. On the other hand,

central editing should take place when all forms or

searching for patterns of relationship that exist among data-groups (Kothari,2004). Thus, "in the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions". But persons like (Selltiz, et.al., 1959) do not like to make difference between processing and analysis. They opine that analysis of data in a general way involves a number of closely related operations which are performed with the purpose of summarizing the collected data and organizing these in such a manner that they answer the research question(s). We, however, shall prefer to observe the difference between the two terms as stated here in order to understand their implications more clearly.

Generally, data analysis in research divided into qualitative and quantitative data analysis. The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis.

schedules have been completed and returned to the office. It implies that all forms should get a thorough editing by a single editor in a small study and by a team of editors in case of a large inquiry. Editor(s) may correct the obvious errors such as an entry in the wrong place, entry recorded in months when it should have been recorded in weeks, and the like.

2. Coding: Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusivity which means that a specific answer can be placed in one and only one cell in a given category set (Kothari ,2004). In addition, Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to pre-code the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires (Neuman, 2000).

3. Classification: Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in

groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

(a) Classification according to attributes: data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as *statistics of attributes* and their classification is said to be classification according to attributes. Such classification can be simple classification or manifold classification (Kothari, 2004).

(b) Classification according to class-intervals: unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as *statistics of variables* and are classified on the basis of class intervals. All the classes or groups, with the irrespective frequencies taken together and put in the form of a table, are described as group frequency distribution or simply frequency distribution. Classification according to class intervals usually involves the following three main problems:

(i) How many classes should be there? What should be their magnitudes?

There can be no specific answer with regard to the number of classes. The decision about this calls for skill and experience of the researcher. However, the objective should be to display the data in such a way as to make it meaningful for the analyst. Typically, we may have 5 to 15 classes. With regard to the second part of the question, we can say that, to the extent possible, class-intervals should be of equal magnitudes, but in some cases unequal magnitudes may result in better classification.

Hence the researcher's objective judgment plays an important part in this connection. Multiples of 2, 5 and 10 are generally preferred while determining class magnitudes. Some statisticians adopt the following formula, suggested by Sturges as cited in Kothari, (2004), determining the size of class interval:

$$i = R / (1 + 3.3 \log N)$$

Where i = size of class interval;

R = Range (i.e., difference between the values of the largest item and smallest item among the given items);
 N = Number of items to be grouped.

It should also be kept in mind that in case one or two or very few items have very high or very low values, one may use what are known as open-ended intervals in the overall frequency distribution.

(ii) How to choose class limits?

While choosing class limits, the researcher must take into consideration the criterion that the mid-point (generally worked out first by taking the sum of the upper limit and lower limit of a class and then divide this sum by 2) of a class-interval and the actual average of items of that class interval should remain as close to each other as possible. Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures. Class limits may generally be stated in any of the following forms:

Exclusive type class intervals: They are usually stated as follows:

10–20 read as above 10 and under 20
 20–30 read as above 10 and under 20
 30–40 read as above 10 and under 20
 40–50 read as above 10 and under 20

Thus, under the exclusive type class intervals, the items whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly 30 would be put in 30–40 class intervals and not in 20–30 class intervals.

In simple words, we can say that under exclusive type class intervals, the upper limit of a class interval is excluded and items with values less than the upper limit (but not less than the lower limit) are put in the given class interval.

Inclusive type class intervals: They are usually stated as follows:

11–20
 21–30
 31–40
 41–50

In inclusive type class intervals the upper limit of a class interval is also included in the concerning class interval. Thus, an item whose value is 20 will be put in 11–20 class intervals. The stated upper limit of the class interval 11–20 is 20 but the real limit is 20.99999 and as such 11–20 class interval really means 11 and under 21.

When the phenomenon under consideration happens to be a discrete one (i.e., can be measured and stated only in integers), then we should adopt inclusive type classification. But when the phenomenon happens to be a continuous one capable of being measured in fractions as well, we can use exclusive type class intervals.

4. Tabulation: When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. This procedure is referred to as tabulation. Thus, tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in columns and rows. As Kothari (2004) stated tabulation is essential because of the following reasons;

- it conserves space and reduces explanatory and descriptive statement to a minimum ,
- it facilitates the process of comparison,
- it facilitates the summation of items and the detection of errors and omissions, and
- it provides a basis for various statistical computations.

Generally, in the process of data analysis the above four steps need to be critical applied. Because, without applying the above process operation of data one cannot do good data analysis.

Qualitative Data Analysis

Concept of Qualitative Data Analysis

Data that is represented either in a verbal or narrative format is qualitative data. These types of data are collected through focus groups, interviews, opened ended questionnaire items, and other less structured situations. A simple way to look at qualitative data is to think of qualitative data in the form of words Migrant & Seasonal Head Start, (2006) stated as:

Qualitative data analysis is the classification and interpretation of linguistic (or visual) material to make statements about implicit and explicit dimensions and structures of meaning-making in the material and what is represented in it. Meaning-making can refer to subjective or social meanings.

From the above explanation we can understand that qualitative data analysis is one way of data analysis which helps to describe or interpret the data through words which transfer information through different

dimensions.

Qualitative data analysis is the range of processes and procedures whereby we move from the qualitative data that have been collected, into some form of explanation, understanding or interpretation of the people and situations we are investigating (Cohen, et.al. 2007). It is usually based on an interpretative philosophy. The idea is to examine the meaningful and symbolic content of qualitative data. It refers to non-numeric information such as interview transcripts, notes, video and audio recordings, images and text documents.

Qualitative data analysis can be divided into the following five categories:

1. **Content analysis:** This refers to the process of categorizing verbal or behavioral data to classify, summarize and tabulate the data. According to Cohen, et.al. (2007), content analysis is the procedure for the categorization of verbal or behavioral data for the purpose of classification, summarization and tabulation. Content analysis can be done on two levels:

- a. **Descriptive:** What is the data? And
 - b. **Interpretative:** what was meant by the data?
2. **Narrative analysis:** This method involves the reformulation of stories presented by respondents taking into account context of each case and different experiences of each respondent. In other words, narrative analysis is the revisions of primary qualitative data by researcher. Narratives are transcribed experiences. Every interview/observation has narrative aspect in which the researcher has to sort-out and reflects up on them, enhance them and present them in a revised shape to the reader. The core activity in narrative analysis is to reformulate stories presented by people in different contexts and based on their different experiences.
3. **Discourse analysis:** A method of analysis of naturally occurring talk and all types of written text. This is a method of analyzing a naturally occurring talk (spoken interaction) and all types of written texts. It focuses on how people express themselves verbally in their everyday social life i.e. how language is used in everyday situations?
- a. Sometimes people express themselves in a simple and straightforward way
 - b. Sometimes people express themselves vaguely and indirectly
 - c. Analyst must refer to the context when interpreting the message because the same phenomenon can be described in a number of different ways depending on context.

4. **Framework analysis:** This is more advanced method that consists of several stages such as **familiarization** (Transcribing & reading the data), **identifying a thematic framework** (Initial coding framework which is developed both from a priori issues and from emergent issues), **coding** (Using numerical or textual codes to identify specific piece of data which correspond to different themes), **charting** (Charts created using headings from thematic framework), and **mapping and interpretation** (Searching for patterns, associations, concepts and explanations in the data).
5. **Grounded theory:** as Corbin and Nicholas (2005) this method of qualitative data analysis starts with an analysis of a single case to formulate a theory. Then, additional cases are examined to see if they contribute to the theory. This theory starts with an examination of a single case from a 'pre-defined' population in order to formulate a general statement (concept or a hypothesis) about a population. Afterwards the analyst examines another case to see whether the hypothesis fits the statement. If it does, a further case is selected but if it doesn't fit there are two options: Either the statement is changed to fit both cases or the definition of the population is changed in such a way that the case is no longer a member of the newly defined population. Then another case is selected and the process continues. In such a way one should be able to arrive at a statement that fits all cases of a population-as-defined. This method is only for limited set of analytic problems: those that can be solved with some general overall statement (Cohen, et.al. 2007).

Aims of Qualitative Data Analysis

The analysis of qualitative data can have several aims. Neuman (2000) explained that: The first aim may be to describe a phenomenon in some or greater detail. The phenomenon can be the subjective experiences of a specific individual or group (e.g. the way people continue to live after a fatal diagnosis). This can focus on the case (individual or group) and its special features and the links between them. The analysis can also focus on comparing several cases (individuals or groups) and on what they have in common or on the differences between them. The second aim may be to identify the conditions on which such differences are based. This means to look for explanations for such differences (e.g. circumstances which make it more likely that the coping with a specific illness situation is more successful than in other cases).

And the third aim may be to develop a theory of the phenomenon under study from the analysis of empirical material (e.g. a theory of illness trajectories).

Advantages and disadvantages of qualitative analysis

Advantages of Qualitative Analysis

Qualitative data analysis has different advantages. Denscombe (2007), stated there are a number of advantages such as:

The first is **the data and the analyses are 'grounded'**. A particular strength associated with qualitative research is that the descriptions and theories such research generates are 'grounded in reality'. This is not to suggest that they depict reality in some simplistic sense, as though social reality were 'out there' waiting to be 'discovered'. But it does suggest that the data and the analysis have their roots in the conditions of social existence. There is little scope for 'armchair theorizing' or 'ideas plucked out of thin air'.

The second, **there is a richness and detail to the data**. The in-depth study of relatively focused areas, the tendency towards small-scale research and the generation of 'thick descriptions' mean that qualitative research scores well in terms of the way it deals with complex social situations. It is better able to deal with the intricacies of a situation and do justice to the subtleties of social life.

The third, **there is tolerance of ambiguity and contradictions**. To the extent that social existence involves uncertainty, accounts of that existence ought to be able to tolerate ambiguities and contradictions, and qualitative research is better able to do this than quantitative research (Maykut and Morehouse, 1994 as cited in Denscombe, 2007). This is not a reflection of a weak analysis. It is a reflection of the social reality being investigated.

Lastly, there **is the prospect of alternative explanations**. Qualitative analysis, because it draws on the interpretive skills of the researcher, opens up the possibility of more than one explanation being valid. Rather than a presumption that there must be, in theory at least, one correct explanation, it allows for the possibility that different researchers might reach different conclusions, despite using broadly the same methods.

A. Disadvantages of Qualitative Analysis

In relation to the disadvantages Denscombe (2007),

also explained the following are some powerful disadvantages:

First, ***the data might be less representative***. The flip-side of qualitative research's attention to thick description and the grounded approach is that it becomes more difficult to establish how far the findings from the detailed, in-depth study of a small number of instances may be generalized to other similar instances. Provided sufficient detail is given about the circumstances of the research, however, it *is* still possible to gauge how far the findings relate to other instances, but such generalizability is still more open to doubt than it is with well conducted quantitative research.

Second, ***interpretation is bound up with the 'self' of the researcher***. Qualitative research recognizes more openly than does quantitative research that the researcher's own identity, background and beliefs have a role in the creation of data and the analysis of data. The research is 'self-aware'. This means that the findings are necessarily more cautious and tentative, because it operates on the basic assumption that the findings are a creation of the researcher rather than a discovery of fact. Although it may be argued that quantitative research is guilty of trying to gloss over the point – which equally well applies – the greater exposure of the intrusion of the 'self' in qualitative research inevitably means more cautious approaches to the findings (Denscombe, 2007).

Third, ***there is a possibility of de-contextualizing the meaning***. In the process of coding and categorizing the field notes, texts or transcripts there is a possibility that the word (or images for that matter) get taken literally *out of context*. The context is an integral part of the qualitative data, and the context refers to both events surrounding the production of the data, and events and words that precede and follow the actual extracted pieces of data that are used to form the units for analysis. There is a very real danger for the researcher that in coding and categorizing of the data the meaning of the data is lost or transformed by wrenching it from its location (a) within a sequence of data (e.g. interview talk), or (b) within surrounding circumstances which have a bearing on the meaning of the unit as it was originally conceived at the time of data collection (Denscombe, 2007).

Fourth, ***there is the danger of oversimplifying the explanation***. In the quest to identify themes in the data and to develop generalizations the researcher can feel pressured to underplay, possibly disregard data that 'doesn't fit'. Inconsistencies, ambiguities and alternative explanations can be frustrating in the way they inhibit a nice clear generalization – but they are an inherent feature of social life. Social phenomena are complex, and the analysis of qualitative data needs to acknowledge this

and avoid attempts to oversimplify matters (Denscombe, 2007).

Fifth, the ***analysis takes longer***. The volume of data that a researcher collects will depend on the time and resources available for the research project. When it comes to the analysis of that data, however, it is almost guaranteed that it will seem like a daunting task (Denscombe, 2007).

Quantitative Data Analysis

Quantitative data is expressed in numerical terms, in which the numeric values could be large or small. Numerical values may correspond to a specific category or label. Quantitative analysis is statistically reliable and generalizable results. In quantitative research we classify features, count them, and even construct more complex statistical models in an attempt to explain what is observed. Findings can be generalized to a larger population, and direct comparisons can be made between two corpora, so long as valid sampling and significance techniques have been used (Bryman and Cramer, 2005). Thus, quantitative analysis allows us to discover which phenomena are likely to be genuine reflections of the behavior of a language or variety, and which are merely chance occurrences. The more basic task of just looking at a single language variety allows one to get a precise picture of the frequency and rarity of particular phenomena, and thus their relative normality or abnormality.

However, the picture of the data which emerges from quantitative analysis is less rich than that obtained from qualitative analysis. For statistical purposes, classifications have to be of the hard-and-fast (so-called "Aristotelian" type). An item either belongs to class x or it doesn't. So in the above example about the phrase "the red flag" we would have to decide whether to classify "red" as "politics" or "color". As can be seen, many linguistic terms and phenomena do not therefore belong to simple, single categories: rather they are more consistent with the recent notion of "fuzzy sets" as in the red example. Quantitative analysis is therefore an idealization of the data in some cases. Also, quantitative analysis tends to sideline rare occurrences. To ensure that certain statistical tests (such as chi-squared) provide reliable results, it is essential that minimum frequencies are obtained - meaning that categories may have to be collapsed into one another resulting in a loss of data richness (Dawson, 2002). So, in generally concept quantitative data analysis is mainly use numbers, graphs, charts, equations, statistics(inferential and descriptive), ANOVA, ANCOVA, regression, and correlation etc.

Statistical Analysis of Data

Statistics is the body of mathematical techniques or processes for gathering, describing organizing and interpreting numerical data. Since research often yields such quantitative data, statistics is a basic tool of measurement and research. The researcher who uses statistics is concerned with more than the manipulation of data, statistical methods goes back to fundamental purposes of analysis. Research in education may deal with two types of statistical data application: Descriptive Statistical Analysis, and Inferential Statistical Analysis. To understand the difference between descriptive and inferential statistics, you must first understand the difference between populations and samples.

A population is the entire collection of a carefully defined set of people, objects, or events(Celine, 2017).

So, population is the broader group of people to whom your results will apply. For example, if the a researcher wants to conduct a research in education (eg. Grade 8 Students language skills in Debre Markos Administration town primary schools) all grade 8 students in that specific area are considered to be population in which the samples will be taken.

A sample is a subset of the people, objects, or events selected from that population (Celine, 2017).

So, sample is the group of individuals who participate in your study. For example, selected grade 8 students from the total population can be the sample for the research.

A. Descriptive Statistics

Descriptive statistics is the type of statistics that probably springs to most people's minds when they hear the word "statistics." In this branch of statistics, the goal is to describe. As Weiss (1999) stated that numerical measures are used to tell about features of a set of data. There are a number of items that belong in this portion of statistics, such as:

The average, or measure of the center of a data set, consisting of the mean, median, mode, or midrange, The spread of a data set, which can be measured with the range or standard deviation, Overall descriptions of data such as the five number summary, Measurements such as skewness and kurtosis.

The exploration of relationships and correlation between paired data, and the presentation of statistical results in graphical form. These measures are important and useful because they allow scientists to see patterns among data, and thus to make

sense of that data. Descriptive statistics consist of methods for organizing and summarizing information (Weiss, 1999)

A *parameter* is a descriptive characteristic of a population (Hinkle, Wiersma, & Jurs, 2003). For example, if we found the average of language skills all grade 8 students mentioned above in the town be it population, the resulting average (also called the *mean*) would be a population parameter. To obtain this average, we first need to tabulate the amount of numerated skills of every student. When calculating this mean, we are engaging in descriptive statistical analysis.

As, Weiss (1999) explained that descriptive *statistical analysis* focuses on the exhaustive measurement of population characteristics. You define a population, assess each member of that population, and compute a summary value (such as a mean or standard deviation) based on those values. It is concerned with numerical description of a particular group observed and any similarity to those outside the group cannot be taken for granted. The data describe one group and that one group only. Much simple educational research involves descriptive statistics and provides valuable information about the nature of a particular group or class.

Data collected from tests and experiments often have little meaning or significance until they have been classified or rearranged in a systematic way. This procedure leads to the organization of materials into few heads:

- (i) Determination of range of the interval between the largest and smallest scores.
- (ii) Decision as to the number and size of the group to be used in classification.

Class interval is therefore, helpful for grouping the data in suitable units and the number and size of these class intervals will depend upon the range of scores and the kinds of measures with which one is dealing. The number of class intervals which a given range will yield can be determined approximately by dividing the range by the interval tentatively chosen.

According to Agresti, & Finlay (1997), the most commonly used methods of analysis data statistically are: Calculating frequency distribution usually in percentages of items under study, testing data for normality of distribution skewness and kurtosis, calculating percentiles and percentile ranks, calculating measures of central tendency-mean, median and mode and establishing norms, calculating measures of dispersion-standard deviation mean deviation, quartile deviation and range, calculating measures of relationship-coefficient of

correlation, reliability and validity by the Rank-difference and Product moment methods, and graphical presentation of data-Frequency polygon curve, Histogram, Cumulative frequency polygon and Ogive etc.

While analyzing data investigations usually make use of as many of the above simple statistical devices as necessary for the purpose of their study. There are two kinds of descriptive statistics that social scientists use:

Measures of central tendency -mean, median, and mode are included under this category. *Measures of central tendency capture general trends within the data and are calculated and expressed as the mean, median, and mode. A mean tells scientists the mathematical average of all of a data set, such as the average age at first marriage; the median represents the middle of the data distribution, like the age that sits in the middle of the range of ages at which people first marry; and, the mode might be the most common age at which people first marry (Huck, 2004).*

The above explanation indicates that the central tendency of a distribution is an estimate of the "center" of a distribution of values. A measure of central tendency is a central or typical value for a probability distribution. Let us see the following examples:

Example one: consider the test score values: 15, 20, 21, 20, 36, 15, 25, 15. The sum of these 8 values is 167, so the mean is $167/8 = 20.875$.

Example two: if there are 500 scores in the list, score 250 would be the median. If we order the 8 scores shown above, we would get: 15,15,15,20,20,21,25,36. There are 8 scores and score 4 and 5 represent the halfway point. Since both of these scores are 20, the median is 20. If the two middle scores had different values, you would have to interpolate to determine the median.

Example three: in a bimodal distribution there are two values that occur most frequently. Notice that for the same set of 8 scores we got three different values -- 20.875, 20, and 15 -- for the mean, median and mode respectively. If the distribution is truly normal (i.e., bell-shaped), the mean, median and mode are all equal to each other.

Measures of spread- variance, standard deviation, quartiles and others are included under this. As Strauss and Corbin (1990) stated that:

Measures of spread describe how the data are distributed and relate to each other, including: The range, the entire range of values present in a data set, The frequency distribution, which defines how many times a particular value occurs within a data set, Quartiles, subgroups formed within a data set when all values are divided into four equal parts across the range, Mean absolute deviation, the average of how much each value deviates from the mean, Variance, which illustrates how much of a spread exists in the data, Standard deviation, which illustrates the spread of data relative to the mean.

So, the above explanation shows that, Measures of spread are often visually represented in tables, pie and bar charts, and histograms to aid in the understanding of the trends within the data. These are ways of summarizing a group of data by describing how spreads out the scores are. It describes how similar or varied the set of observed values are for a particular variable (data item). For example, the mean score of our 100 students may be 65 out of 100. However, not all students will have scored 65 marks. Rather, their scores will be spread out. Some will be lower and others higher. Measures of spread help us to summarize how spreads out these scores are. To describe this spread, a number of statistics are available to us, including the range, quartiles, absolute deviation, variance and standard deviation.

Generally, Descriptive statistics includes the construction of graphs, charts, and tables, and the calculation of various descriptive measures such as averages, measures of variation, and percentiles. In fact, the most part of this course deals with descriptive statistics.

B. Inferential Statistics

The second type of statistics is inferential statistics. Inferential statistical analysis involves the process of sampling, the selection for study of a small group that is assumed to be related to the large group from which it is drawn. Agresti & Finlay (1997) stated the small group is known as the sample; the large group, the population or universe. A statistics is a measure based on a sample. A statistic computed from a sample may be used to estimate a parameter, the corresponding value in the population which it is selected. This is a set of methods used to make a generalization, estimate, prediction or decision. Inferential statistics is the mathematics and

logic of how this generalization from sample to population can be made. The fundamental question is: can we infer the population's characteristics from the sample's characteristics?

Ex. Of 350 (grade 8) randomly selected students in the town of Debre Markos out of 2000 students and their average listening skills tests is calculated to be (75%), this is sample result that we can make generalization about the total population (2000 students) which is inferential statistics.

The major use of inferential statistics is to use information from a **sample** to infer something about a **population**. Inferential statistics consist of methods for drawing and measuring the reliability of conclusions about population based on information obtained from a sample of the population (Weiss, 1999).

Inferential statistics are produced through complex mathematical calculations that allow scientists to infer trends about a larger population based on a study of a sample taken from it. Scientists use inferential statistics to examine the relationships between variables within a sample and then make generalizations or predictions about how those variables will relate to a larger population.

A measured value based upon sample data is statistic. A population value estimated from a statistic is a parameter. A sample is a small proportion of a population selected for analysis. By observing the sample, certain inferences may be made about the population. Samples are not selected haphazardly, but are chosen in a deliberate way so that the influence of chance or probability can be estimated. The basic ideas of inference are to estimate the parameters with the help of sample statistics which play an extremely important role in educational research. These basic ideals, of which the concept of underlying distribution is a part, comprise the foundation for testing hypotheses using statistical techniques.

The parameters are never known for certain unless the entire population is measured and then there is no inference. We look at the statistics and their underlying distributions and from them we reason to tenable conclusions about the parameters.

It is usually impossible to examine each member of the population individually. So scientists choose a representative subset of the population, called a statistical sample, and from this analysis, they are able to say something about the population from which the sample came. There are two major divisions of inferential statistics (Agresti, & Finlay, (1997) stated as follow:

A confidence interval gives a range of values for an unknown parameter of the population by measuring a statistical sample. This is expressed in terms of an interval and the degree of confidence that the parameter is within

the interval.

Tests of significance or hypothesis testing where scientists make a claim about the population by analyzing a statistical sample. By design, there is some uncertainty in this process. This can be expressed in terms of a level of significance.

The above explanation shows us, in statistics, a confidence interval (CI) is a type of interval estimate (of a population parameter) that is computed from the observed data. The confidence level is the frequency (i.e., the proportion) of possible confidence intervals that contain the true value of their corresponding parameter. And once sample data has been gathered through an observational study or experiment, statistical inference allows analysts to assess evidence in favor or some claim about the population from which the sample has been drawn. The methods of inference used to support or reject claims based on sample data are known as tests of significance.

Furthermore, Howell, (2002) stated that a statistic is a numerical value that is computed from a sample, describes some characteristic of that sample such as the mean, and can be used to make inferences about the population from which the sample is drawn. For example, if you were to compute the average amount of insurance sold by your sample of 100 agents, that average would be a statistic because it summarizes a specific characteristic of the sample. Remember that the word "statistic" is generally associated with samples, while "parameter" is generally associated with populations. In similar taken, Weiss, (1999) in contrast to descriptive statistics, inferential statistical analysis involves using information from a sample to make inferences, or estimates, about the population.

Inferential statistics consist of methods for drawing and measuring the reliability of conclusions about population based on information obtained from a sample of the population (Weiss, 1999). In short, inferential statistics includes methods like point estimation, interval estimation and hypothesis testing which are all based on probability theory.

Example (Descriptive and Inferential Statistics).

Consider event of tossing dice. The dice is rolled 100 times and the results are forming the sample data. Descriptive statistics is used to grouping the sample data to the following table.

Outcome of the roll Frequencies in the sample data

Outcome of the roll	Frequencies in the sample data
1	10
2	20
3	18
4	16
5	11
6	25

Inferential statistics can now be used to verify whether the dice is a fair or not.

Generally, descriptive and inferential statistics are interrelated. It is almost always necessary to use methods of descriptive statistics to organize and summarize the information obtained from a sample before methods of inferential statistics can be used to make more thorough analysis of the subject under investigation. Furthermore, the preliminary descriptive analysis of a sample often reveals features that lead to the choice of the appropriate inferential method to be later used. Sometimes it is possible to collect the data from the whole population. In that case it is possible to perform a descriptive study on the population as well as usually on the sample. Only when an inference is made about the population based on information obtained from the sample does the study become inferential.

Analysis of Variance (ANOVA)

Concept of ANOVA

One of the methods for quantitative data analysis is analysis of variance. According to Kothari (2004), Professor R.A. Fisher was the first man to use the term 'Variance' and, in fact, it was he who developed a very elaborate theory concerning ANOVA, explaining its usefulness in practical field. Later on Professor Snedecor and many others contributed to the development of this technique.

ANOVA is essentially a procedure for testing the difference among different groups of data for homogeneity. "The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to chance and that amount which can be attributed to specified causes "(Bryman, and Cramer, 2005). There may be variation between samples and also within sample items.

Cramer, (2005) stated that the specific analysis of variance test that we will study is often referred to as the one way ANOVA. It consists in splitting the variance for analytical purposes. Hence, it is a method of analyzing the variance to which a response is subject into its

various components corresponding to various sources of variation. Through this technique one can explain whether various varieties of seeds or fertilizers or soils differ significantly so that a policy decision could be taken accordingly, concerning a particular variety in the context of agriculture researches (Cramer, 2005). Similarly, the differences in various types of feed prepared for a particular class of animal or various types of drugs manufactured for curing a specific disease may be studied and judged to be significant or not through the application of ANOVA technique. Likewise, a manager of a big concern can analyze the performance of various salesmen of his concern in order to know whether their performances differ significantly (Neuman, 2006). ANOVA can be one way or two way ANOVA.

One way (single –factor) ANOVA- Under the one-way ANOVA, we consider only one factor and then observe that the reason for said factor to be important is that several possible types of samples can occur within that factor (Neuman, 2006); Armstrong, Eperjesi and Gilmartin, 2002). We then determine if there are differences within that factor.

Two-Way ANOVA- this technique is used when the data are classified on the basis of two factors. For example, the agricultural output may be classified on the basis of different varieties of seeds and also on the basis of different varieties of fertilizers used. A business firm may have its sales data classified on the basis of different salesmen and also on the basis of sales in different regions. In a factory, the various units of a product produced during a certain period may be classified on the basis of different varieties of machines used and also on the basis of different grades of labour (Neuman, 2006);. Such a two-way design may have repeated measurements of each factor or may not have repeated values.

Assumptions of ANOVA

Like so many of our inference procedures, ANOVA has some underlying assumptions which should be in place in

order to make the results of calculations completely trustworthy. In relation to the assumption of ANOVA, Huck (2004) stated that: ***Subjects are chosen via a simple random sample, Within each group/population, the response variable is normally distributed, While the population means may be different from one group to the next, the population standard deviation is the same for all groups.***

Fortunately, ANOVA is somewhat robust (i.e., results remain fairly trustworthy despite mild violations of these assumptions). Assumptions (ii) and (iii) are close enough to being true if, after gathering ***simple random samples*** from each group, we: Look at normal quintile plots for each group and, in each case, see that the data points fall close to a line, and Compute the standard deviations for each group sample, and see that the ratio of the largest to the smallest group sample standard deviation is no more than two(Cramer, (2005); Neuman, (2006) and Armstrong, Eperjesi and Gilmartin, 2002).

Uses of ANOVA

The one-way analysis of variance for independent groups applies to an experimental situation where there might be more than two groups. The t-test was limited to two groups, but the ANOVA can analyze as many groups as you want. Examine the relationship between variables when there is a nominal level independent variable has 3 or more categories and a normally distributed interval/ratio level of dependent variable produces an F-ratio, which determines the statistical significance of the result. Reduces the probability of a Type I error (which would occur if we did multiple t-tests rather than one single ANOVA)(Singh, 2007).

In relation to the use of ANOVA, Mordkoff (2016) stated that One-way ANOVA is used to test for significant differences among sample means, differs from t-test since more than 2 groups are tested, simultaneously, one factor (independent variable) is analyzed, also called the "grouping" variable, and dependent variable should be interval or ratio but independent variable is usually nominal.

A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one independent variable affecting a dependent variable. With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one measurement variable (i.e. a quantitative variable) and two nominal variables. In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate. Assumptions for Two Way ANOVA: The population must be close to a normal

distribution, Samples must be independent, Population variances must be equal, and Groups must have equal sample sizes (Mordkoff (2016).

Generally, An ANOVA tests whether one or more samples means are significantly different from each other. To determine which or how many sample means are different requires post hoc testing. Two samples where means are significantly different. These two sample means are NOT significantly different due to smaller difference and high variability. Even with same difference between means, if variances are reduced the means can be significantly different.

Analysis of Co-variance (ANCOVA)

The Analysis of Covariance (generally known as ANCOVA) is a technique that sits between analysis of variance and regression analysis. It has a number of purposes but the two that are, perhaps, of most importance are: to increase the precision of comparisons between groups by accounting to variation on important prognostic variables, and to "adjust" comparisons between groups for imbalances in important prognostic variables between these groups.

When we measure covariates and include them in an analysis of variance we call it analysis of covariance (or ANCOVA for short). There are two reasons for including covariates in ANOVA:

1. To reduce within-group error variance: *In the discussion of ANOVA and t-tests we got used to the idea that we assess the effect of an experiment by comparing the amount of variability in the data that the experiment can explain against the variability that it cannot explain. If we can explain some of this 'unexplained' variance (SSR) in terms of other variables (covariates), then we reduce the error variance, allowing us to more accurately assess the effect of the independent variable (SSM) (Hinkle, Wiersma, &Jurs, 2003).*

2. Elimination of confounds: *In any experiment, there may be unmeasured variables that confound the results (i.e. variables that vary systematically with the experimental manipulation). If any variables are known to influence the dependent variable being measured, then ANCOVA is ideally suited to remove the bias of these variables. Once a possible confounding variable has been identified, it can be measured and entered into the analysis as a covariate Hinkle, Wiersma, &Jurs, 2003).*

The above two explanations indicates that the reason for including covariates is that covariates are a variable that a researcher seeks to control for (statistically

subtract the effects of) by using such techniques as multiple regression analysis (MRA) or analysis of covariance (ANCOVA).

But, there are other reasons for including covariates in ANOVA but because I do not intend to describe the computation of ANCOVA in any detail I recommend that the interested reader consult my favorite sources on the topic (Stevens, 2002; Wildt & Ahtola, 1978). Imagine that the researcher who conducted the Viagra study in the previous chapter suddenly realized that the libido of the participants' sexual partners would affect the participants' own libido (especially because the measure of libido was behavioral). Therefore, they repeated the study on a different set of participants, but this time took a measure of the partner's libido. The partner's libido was measured in terms of how often they tried to initiate sexual contact.

Analysis of Covariance (ANCOVA) is an extension of ANOVA that provides a way of statistically controlling the (linear) effect of variables one does not want to examine in a study. These extraneous variables are called covariates, or control variables. (Covariates should be measured on an interval or ratio scale.) (Vogt, 1999). It allows you to remove covariates from the list of possible explanations of variance in the dependent variable. ANCOVA does this by using statistical techniques (such as regression to partial out the effects of covariates) rather than direct experimental methods to control extraneous variables. ANCOVA is used in experimental studies when researchers want to remove the effects of some antecedent variable. For example, pretest scores are used as covariates in pretest posttest experimental designs. ANCOVA is also used in non-experimental research, such as surveys or nonrandom samples, or in quasi-experiments when subjects cannot be assigned randomly to control and experimental groups. Although fairly common, the use of ANCOVA for non-experimental research is controversial (Vogt, 1999).

Assumptions and Issues in ANCOVA

In addition to the assumptions underlying the ANOVA, there are two major assumptions that underlie the use of ANCOVA; both concern the nature of the relationship between the dependent variable and the covariate (Howell, 2002; Huck, 2004; and Vogt, 1999). They stated the assumptions as follow:

The first is that the relationship is linear. If the relationship is nonlinear, the adjustments made in the ANCOVA will be biased; the magnitude of this bias depends on the degree of departure from linearity, especially when there are substantial differences between the groups on the covariate. Thus it is important

for the researcher, in preliminary analyses, to investigate the nature of the relationship between the dependent variable and the covariate (by looking at a scatter plot of the data points), in addition to conducting an ANOVA on the covariate (Howell (2002); Huck (2004) and Vogt, 1999).

The second assumption has to do with the regression lines within each of the groups, (Howell (2002); Huck, S. W. (2004) and Vogt, W. P. (1999). We assume the relationship to be linear. Additionally, however, the regression lines for these individual groups are assumed to be parallel; in other words, they have the same slope. This assumption is often called **homogeneity of regression slopes** or **parallelism** and is necessary in order to use the pooled within-groups regression coefficient for adjusting the sample means and is one of the most important assumptions for the ANCOVA.

Failure to meet this assumption implies that there is an interaction between the covariate and the treatment. This assumption can be checked with an F test on the interaction of the independent variable(s) with the covariate(s). If the F test is significant (i.e., significant interaction) then this assumption has been violated and the covariate should not be used as is. A possible solution is converting the continuous scale of the covariate to a categorical (discrete) variable and making it a subsequent independent variable, and then use a factorial ANOVA to analyze the data.

Moreover, the assumptions underlying the ANCOVA had a slight modification from those for the ANOVA, however, conceptually, they are the same. According to Hinkle, Wiersma, & Jurs, (2003) ANCOVA has the following assumptions:

Assumption 1: The cases represent a random sample from the population, and the scores on the dependent variable are independent of each other, known as the **assumption of independence**.

The test will yield inaccurate results if the independence assumption is violated. This is a design issue that should be addressed prior to data collection. Using random sampling is the best way of ensuring that the observations are independent; however, this is not always possible. The most important thing to avoid is having known relationships among participants in the study.

Assumption 2: The dependent variable is normally distributed in the population for any specific value of the covariate and for any one level of a factor (independent variable), known as the **assumption of normality**.

This assumption describes multiple conditional distributions of the dependent variable, one for every combination of values of the covariate and levels of the factor, and requires them all to be normally distributed. To the extent that population distributions are not normal and sample sizes are small, p values may be invalid. In addition, the power of ANCOVA tests may be reduced considerably if the population distributions are non-normal and, more specifically, thick-tailed or heavily skewed. The assumption of normality can be checked with skewness values (e.g., within +3.29 standard deviations).

Assumption 3: The variances of the dependent variable for the conditional distributions are equal, known as the assumption of homogeneity of variance. To the extent that this assumption is violated and the group sample sizes differ, the validity of the results of the one-way ANCOVA should be questioned. Even with equal sample sizes, the results of the standard post hoc tests should be mistrusted if the population variances differ. The assumption of homogeneity of variance can be checked with the Levine's F-test.

It happens when they're trying to run an analysis of covariance (ANCOVA) model because they have a categorical independent variable and a continuous covariate. The problem arises when a coauthor, committee member, or reviewer insists that ANCOVA is inappropriate in this situation because one of the following ANCOVA assumptions is not met: The independent variable and the covariate are independent of each other, and there is no interaction between independent variable and the covariate (Helwig, 2017).

Regression and Correlation

A. Regression

Regression analysis is used in statistics to find trends in data. For example, we might guess that there's a connection between how much we eat and how much we weigh; regression analysis can help us quantify that. Regression analysis will provide us with an equation for a graph so that we can make predictions about our data. For example, if we've been putting on weight over the last few years, it can predict how much we'll weigh in ten years' time if we continue to put on weight at the same rate. It will also give us a slew of statistics (including a p-value and a correlation coefficient) to tell us how accurate your model is. Most elementary statistics courses cover very basic techniques, like making scatter plots and performing linear regression. However, we may come across more advanced techniques like multiple regressions Gogtay, Deshpande ,and Thatte 2017).

Furthermore, Huck (2004) stated that Regression analysis is a way of predicting an **outcome variable** from one **predictor variable**(**simple regression**) or several predictor variables (**multiple regressions**).

This tool is incredibly useful because it allows us to go a step beyond the data that we collected. So, this indicates that Regression analysis is a statistical technique for investigating the relationship among variables. O'Brien, and Scott (2012) stated that the concept of regression as follow:

Regression is particularly useful to understand the predictive power of the independent variables on the dependent variable once a causal relationship has been confirmed. To be precise, regression helps a researcher understand to what extent the change of the value of the dependent variable causes the change in the value of the independent variables, while other independent variables are held unchanged (p. 3).

From the above explanation we can understand that, regression is one of a tool for quantitative analysis which is used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships used to infer causal relationships between the independent and dependent variables.

In regression analysis, the problem of interest is the nature of the relationship itself between the dependent variable (response) and the (explanatory) independent variable. The regression equation describes the relationship between two variables and is given by the general format:

$$FY = a + bX + \epsilon$$

Where: Y = dependent variable;
X = independent variable,
a = intercept of regression line;
b = slope of regression line, and
 ϵ = error term

In this format, given that Y is dependent on X, the slope b indicates the unit changes in Y for every unit change in X. If b = 0.66, it means that every time X increases (or decreases) by a certain amount, Y increases (or decreases) by 0.66 that amount. The intercept a indicates the value of Y at the point where X = 0. Thus if X indicated market returns, the intercept would show how the dependent variable performs when the market has a flat quarter where returns are 0. In investment parlance, a manager has a positive alpha because a linear

regression between the manager's performance and the performance of the market has an intercept number a greater than

Assumptions for regression: there are assumptions to be taken in to consideration to use regression as a tool of data analysis. Gogtay, Deshpande, and Thatte (2017) stated the following assumptions:

Assumption 1: The relationship between the *independent variables and the dependent variables* is linear. The first assumption of Multiple Regression is that the relationship between the IVs and the DV can be characterized by a straight line. A simple way to check this is by producing scatter plots of the relationship between each of our IVs and our DV.

Assumption 2: There is no multi co linearity in your data. This is essentially the assumption that your predictors are not too highly correlated with one another.

Assumption 3: The values of the residuals are independent. This is basically the same as saying that we need our observations (or individual data points) to be independent from one another (or uncorrelated). We can test this assumption using the **Durbin-Watson** statistic.

Assumption 4: The variance of the residuals is constant. This is called *homoscedasticity*, and is the assumption that the variation in the residuals (or amount of error in the model) is similar at each point across the model. In other words, the spread of the residuals should be fairly constant at each point of the predictor variables (or across the linear model). We can get an idea of this by looking at our original scatter plot, but to properly test this, we need to ask SPSS to produce a special scatter plot for us that includes the whole model (and not just the individual predictors).

To test the 4th assumption, we need to plot the standardized values our model would predict, against the standardized residuals obtained.

Assumption 5: The values of the residuals are normally distributed. This assumption can be tested by looking at the distribution of residuals.

Assumption 6: There are no influential cases biasing your model. Significant outliers and influential data points can place undue influence on your model, making it less representative of your data as a whole.

B. Correlation

Correlation is a measure of association between two variables. The variables are not designated as dependent or independent. As O'Brien and Scott (2012), explained

that the two most popular correlation coefficients are: Spearman's correlation coefficient rho and Pearson's product-moment correlation coefficient. When calculating a correlation coefficient for ordinal data, select Spearman's technique. For interval or ratio-type data, use Pearson's technique.

Ott, (1993) stated that:

Correlation is a measure of the strength of a relationship between two variables. Correlations do not indicate causality and are not used to make predictions; rather they help identify how strongly and in what direction two variables co-vary in an environment.

So, from the above definition we can deduce that Correlation analysis is useful when researchers are attempting to establish if a relationship exists between two variables. The correlation coefficient is a measure of the degree of linear association between two continuous variables.

Pearson r correlation: Pearson r correlation is widely used in statistics to measure the degree of the relationship between linear related variables (Gogtay, Deshpande, and Thatte, 2017). For example, in the stock market, if we want to measure how two commodities are related to each other, Pearson r correlation is used to measure the degree of relationship between the two commodities.

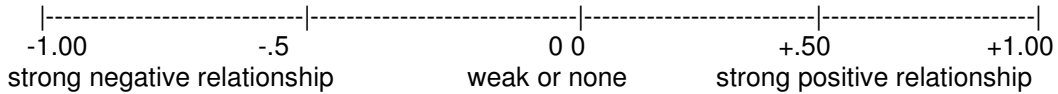
Assumption For the Pearson r correlation, both variables should be normally distributed. Other assumptions include linearity and homoscedasticity. Linearity assumes a straight line relationship between each of the variables in the analysis and homoscedasticity assumes that data is normally distributed about the regression line (Gogtay, Deshpande, and Thatte, 2017).

Spearman rank correlation: Spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. It was developed by Spearman, thus it is called the Spearman rank correlation (Gogtay, Deshpande, and Thatte ,2017). Spearman rank correlation test does not assume any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.

Assumptions: Spearman rank correlation test does not make any assumptions about the distribution. The assumptions of Spearman rho correlation are that data must be at least ordinal and scores on one variable must be monotonically related to the other variable.

The value of a correlation coefficient can vary from **-1 to 1**. -1 indicates a perfect negative correlation, while a +1 indicates a perfect positive correlation. A correlation of **zero** means there is no relationship between the two variables. When there is a negative correlation between two variables, as the value of one variable increases, the

value of the other variable decreases, and vice versa (Gogtay, Deshpande, and Thatte ,2017). In other words, for a negative correlation, the variables work opposite each other. When there is a positive correlation between two variables, as the value of one variable increases, the value of the other variable also increases. The variables move together.



The standard error of a correlation coefficient is used to determine the confidence intervals around a true correlation of zero. If your correlation coefficient falls outside of this range, then it is significantly different than zero. The standard error can be calculated for interval or ratio-type data (i.e., only for Pearson's product-moment correlation).

Example: A company wanted to know if there is a significant relationship between the total number of salespeople and the total number of sales. They collect data for five months.

Variable 1	Variable 2
207	6907
180	5991
220	6810
205	6553
190	6190

Correlation coefficient = .921
 Standard error of the coefficient = .068
 t-test for the significance of the coefficient = 4.100
 Degrees of freedom = 3
 Two-tailed probability = .0263

Generally, as Hinkle, Wiersma, &Jurs, (2003) stated that the goal of a correlation analysis is to see whether two measurement variables co vary, and to quantify the strength of the relationship between the variables, whereas regression expresses the relationship in the form of an equation.

Types of Regression

Gogtay, Deshpande, and Thatte (2017) stated that essentially in research, there are three common types of regression analyses that are used viz., linear, logistic regression and Cox regression. These are chosen

depending on the type of variables that we are dealing with. Cox regression is a special type of regression analysis that is applied to survival or "time to event "data and will be discussed in detail in the next article in the series. Linear regression can be simple linear or multiple linear regressions while Logistic regression could be Polynomial in certain cases.

The type of regression analysis to be used in a given situation is primarily driven by the following three metrics: Number and nature of independent variable/s , Number and nature of the dependent variable/s, and Shape of the regression line.

A. Linear regression: Linear regression is the most basic and commonly used regression technique and is of two type's viz. simple and multiple regressions. You can use Simple linear regression when there is a single dependent and a single independent variable. Both the variables must be continuous and the line describing the relationship is a straight line (linear). Multiple linear regression on the other hand can be used when we have one continuous dependent variable and two or more independent variables. Importantly, the independent variables could be quantitative or qualitative (O'Brien and Scott, 2012). They added that, in simple linear regression, the outcome or dependent variable Y is predicted by only one independent or predictive variable. Multiple regression is not just a technique on its own. It is, in fact, a family of techniques that can be used to explore the relationship between one continuous dependent variable and a number of independent variables or predictors. Although multiple regression is based on correlation, it enables a more sophisticated exploration of the interrelationships among variables.

Both the independent variables here could be expressed either as continuous data or qualitative data. A linear relationship should exist between the dependent and independent variables.

B. Logistic regression: This type of regression analysis is used when the dependent variable is binary in nature. For example, if the outcome of interest is death in a cancer study, any patient in the study can have only one of two possible outcomes- dead or alive. The impact of one or more predictor variables on this binary variable is assessed. The predictor variables can be either quantitative or qualitative. Unlike linear regression, this type of regression does not require a linear relationship between the predictor and dependent variables.

For logistic regression to be meaningful, the following criteria must be met/satisfied: The independent variables must not be correlated amongst each other and the sample size should be adequate. If the dependent variable is non-binary and has more than two possibilities, we use the multinomial or polynomial logistic regression.

Table 1: Types of regression adopted from Gogtay, Deshpande, and Thatte (2017)
The following table will summarize the basic types of regression:

Type of regression	Dependent variable and its nature	Independent variable and its nature	Relationship between variables
Simple linear	One, continuous, normally distributed	One, continuous, normally distributed	Linear
Multiple linear	One, continuous	Two or more, may be continuous or categorical	Linear
Logistic	One, binary	Two or more, may be continuous or categorical	Need not be linear
Polynomial (logistic) [multinomial]	Non-binary	Two or more, may be continuous or categorical	Need not be linear
Cox or proportional hazards regression	Time to an event	Two or more, may be continuous or categorical	Is rarely linear

Multiple Correlation and Regression

When there are two or more than two independent variables, the analysis concerning relationship is known as multiple correlations and the equation describing such relationship as the multiple regression equation. We here explain multiple correlation and regression taking only two independent variables and one dependent variable (Convenient computer programs exist for dealing with a great number of variables).

In correlation, the two variables are treated as equals. In regression, one variable is considered independent (=predictor) variable (X) and the other the dependent (=outcome) variable Y (Quirk, 2007). Prediction: If you know something about X, this knowledge helps you predict something about Y.

In simple linear regression, the outcome or dependent variable Y is predicted by only one independent or predictive variable. It should be stressed that in very rare

cases, the dependent variable can only be explained by one independent variable.

Assumptions behind Multiple Regressions

Multiple regressions make a number of assumptions about the data, and are important that these are met. The assumptions are: Sample Size, Multi co linearity of IVs, Linearity, Absence of outliers, Homo scedasticit, and Normality. so, Tests of these assumptions are numerous so we will only look at a few of the more important ones.

a. Sample size: You will encounter a number of recommendations for a suitable sample size for multiple regression analysis (Tabachnick & Fidell, 2007). As a simple rule, you can calculate the following two values: $104 + m$

$50 + 8m$ where m is the number of independent variables, and take whichever is the largest as the minimum number of cases required.

For example, with 4 independent variables, we would

require at least 108 cases: $[104+4=108]$ $[50+8*4=82]$
 With 8 independent variables we would require at least 114 cases: $[104+8=112]$ $[50+8*8=114]$ With Stepwise regression, we need at least 40 cases for every independent variable (Pallant, 2007). However, when any of the following assumptions is violated, larger samples are required.

b. Multicollinearity of Independent Variables: Any two independent variables with a Pearson correlation coefficient greater than .9 between them will cause problems. Remove independent variables with a tolerance value less than 0.1. A tolerance value is calculated as $1 - R^2$, which is reported in SPSS (Tabachnick & Fidell, 2007).

c. Linearity: Standard multiple regressions only looks at linear relationships. You can check this roughly using bivariate scatterplots of the dependent variable and each of the independent variables (Tabachnick & Fidell, 2007).

d. Absence of outliers: Outliers, such as extreme cases can have a very strong effect on a regression equation. They can be spotted on scatter plots in early stages of your analysis. There are also a number of more advanced techniques for identifying problematic points. These are very important in multiple regression analysis where you are not only interested in extreme values but in unusual combinations of independent values.

e. Homoscedasticity: This assumption is similar to the assumption of homogeneity of variance with ANOVAs. More advanced methods include examining residuals. It requires that there be equality of variance in the independent variables for each value of the dependent variable. We can do this in a crude way with the scatter plots for each independent variable against the dependent variable (Tabachnick & Fidell, 2007). If there is equality of variance, then the points of the scatter plot should form an evenly balanced cylinder around the regression line.

f. Normality: The dependent and independent variables should be normally distributed.

When we talk about Multiple Regressions it can be: Standard Multiple Regressions (All of the independent (or predictor) variables are entered into the equation simultaneously), Hierarchical Multiple Regressions (The independent variables are entered into the equation in the order specified by the researcher based on their theoretical approach), and Stepwise Multiple Regression (The researcher provides SPSS with a list of independent variables and then allows the program to select which

variables it will use and in which order they can go into the equation, based on statistical criteria.

Uses of Correlation and Regression

There are three main uses for correlation and regression. Cohen, (1988) stated that,

- One is to test hypotheses about cause-and-effect relationships. In this case, the experimenter determines the values of the X-variable and sees whether variation in X causes variation in Y. For example, giving people different amounts of a drug and measuring their blood pressure.
- The second main use for correlation and regression is to see whether two variables are associated, without necessarily inferring a cause-and-effect relationship. In this case, neither variable is determined by the experimenter; both are naturally variable. If an association is found, the inference is that variation in X may cause variation in Y, or variation in Y may cause variation in X, or variation in some other factor may affect both X and Y.
- The third common use of linear regression is estimating the value of one variable corresponding to a particular value of the other variable.

Advantages and Disadvantages of Quantitative Analysis

Advantages of Quantitative Analysis

Denscombe, (2007) stated the following advantages of quantitative analysis:

First, it is **Scientific**: Quantitative data lend themselves to various forms of statistical techniques based on the principles of mathematics and probability. Such statistics provide the analyses with an aura of scientific respectability. The analyses *appear* to be based on objective laws rather than the values of the researcher. Second, **Confidence**: Statistical tests of significance give researchers additional credibility in terms of the interpretations they make and the confidence they have in their findings. Third, **Measurement**: The analysis of quantitative data provides a solid foundation for description and analysis. Interpretations and findings are based on measured quantities rather than impressions, and these are, at least in principle, quantities that can be checked by others for authenticity. Fourth, **Analysis**. Large volumes of quantitative data can be analyzed relatively quickly, provided adequate preparation and planning has occurred in advance. Once the procedures are 'up and running', researchers can interrogate their

results relatively quickly. Fifth, **Presentation**. Tables and charts provide a succinct and effective way of organizing quantitative data and communicating the findings to others. Widely available computer software aids the design of tables and charts, and takes most of the hard labor out of statistical analysis.

Disadvantages of Quantitative Analysis

According to Denscombe (2007) the following are some limitations of quantitative data analysis.

First, **quality of data**: The quantitative data are only as good as the methods used to collect them and the questions that are asked. As with computers, it is a matter of 'garbage in, garbage out'. Second, **Technicist**: There is a danger of researchers becoming obsessed with the techniques of analysis at the expense of the broader issues underlying the research. Particularly with the power of computers at researchers' fingertips, attention can sway from the real purpose of the research towards an overbearing concern with the technical aspects of analysis. Third, **Data overload**: Large volumes of data can be strength of quantitative analysis but, without care, it can start to overload the researcher. Too many cases, too many variables, too many factors to consider – the analysis can be driven towards too much complexity. The researcher can get swamped. Fourth, **false promise**: Decisions made during the analysis of quantitative data can have far-reaching effects on the kinds of findings that emerge. In fact, the analysis of quantitative data, in some respects, is no more neutral or objective than the analysis of qualitative data. For example, the manipulation of categories and the boundaries of grouped frequencies can be used to achieve a data fix, to show significance where other combinations of the data do not. *Quantitative analysis is not as scientifically objective as it might seem on the surface.*

DISCUSSION OF RESULTS

Qualitative Data Result

Research Gateway shows us how to discuss the results that we have found in relation to both our research questions and existing knowledge. This is our opportunity to highlight how our research reflects, differs from and extends current knowledge of the area in which we have chosen to carry out research. This section is our chance to demonstrate exactly what we know about this topic by interpreting our findings and outlining what they mean. At the end of our discussion we should have discussed all of the results that we found and provided an explanation for our findings.

Discussion section should not be simply a summary of the results we have found and at this stage we will have to demonstrate original thinking. First, we should highlight and discuss how our research has reinforced what is already known about the area. Many students make the mistake of thinking that they should have found something new; in fact, very few research projects have findings that are unique. Instead, we are likely to have a number of findings that reinforce what is already known about the field and we need to highlight these, explaining why we think this has occurred.

Second, we may have discovered something different and if this is the case, we will have plenty to discuss. We should outline what is new and how this compares to what is already known. We should also attempt to provide an explanation as to why our research identified these differences. Third, we need to consider how our results extend knowledge about the field. Even if we found similarities between our results and the existing work of others, our research extends knowledge of the area, by reinforcing current thinking. We should state how it does this as this is a legitimate finding. It is important that this section is comprehensive and well structured; making clear links back to the literature we reviewed earlier in the project. This will allow us the opportunity to demonstrate the value of our research and it is therefore very important to discuss our work thoroughly.

The resources in this section of the gateway should help us to:

Interpret the research: the key to a good discussion is a clear understanding of what the research means. This can only be done if the results are interpreted correctly.

Discuss coherently: a good discussion presents a coherent, well-structured explanation that accounts for the findings of the research, making links between the evidence obtained and existing knowledge.

As always, use the Gateway resources appropriately. As usual, the resources have been included because we believe they provide accessible, practical and helpful information on how to discuss our work. On the other hand, don't forget that our institution will have requirements of us and our project that override any information that we get from this Gateway. For example, we might not have to produce a separate discussion section (it depends on different institutions and research types) as this may need to be included with the presentation of results. This is often the case for qualitative research, so we must be sure what is needed. Find out, and then use the Gateway accordingly.

When crafting our findings, the first thing we want to think about is how we will organize our findings. Our

findings represent the story we are going to tell in response to the research questions we have answered. Thus, we will want to organize that story in a way that makes sense to us and will make sense to our reader. We want to think about how we will present the findings so that they are compelling and responsive to the research question(s) we answered. These questions may not be the questions we set out to answer but they will definitely be the questions we answered. We may discover that the best way to organize the findings is first by research question and second by theme. There may be other formats that are better for telling our story. Once we have decided how we want to organize the findings, we will start the chapter by reminding our reader of the research questions. We will need to differentiate between presenting raw data and using data as evidence or examples to support the findings we have identified (Cohen et.al.,2007). Here are some points to consider: Our findings should provide sufficient evidence from our data to support the conclusions we have made. Evidence takes the form of quotations from interviews and excerpts from observations and documents, ethically we have to make sure we have confidence in our findings and account for counter-evidence (evidence that contradicts our primary finding) and not report something that does not have sufficient evidence to back it up, our findings should be related back to our conceptual framework, our findings should be in response to the problem presented (as defined by the research questions) and should be the "solution" or "answer" to those questions, and We should focus on data that enables us to answer your research questions, not simply on offering raw data (Neuman, 2000).

Qualitative research presents "best examples" of raw data to demonstrate an analytic point, not simply to display data. Numbers (descriptive statistics) help our reader understand how prevalent or typical a finding is. Numbers are helpful and should not be avoided simply because this is a qualitative dissertation.

Quantitative Data Result

Quantitative data result is on type of result in research which is presented in quantitative way like numbers, statistics. As Creswell, (2013); Neuman, and Robson, (2004); and Neuman, and Neuman, (2006) stated that Quantitative methods are used to examine the relationship between variables with the primary goal being to analyze and represent that relationship mathematically through statistical analysis. This is the type of research approach most commonly used in scientific research problems.

The finding of your study should be written objectively and in a succinct and precise format. In quantitative studies, it is common to use graphs, tables, charts, and

other non-textual elements to help the reader understand the data. Make sure that non-textual elements do not stand in isolation from the text but are being used to supplement the overall description of the results and to help clarify key points being made (Agresti, and Finlay, 1997). Further information about how to effectively present data using charts and graphs can be found here.

Quantitative Research is used to quantify the problem by way of generating numerical data or data that can be transformed into usable statistics. It is used to quantify attitudes, opinions, behaviors, and other defined variables and generalize results from a larger sample population. Quantitative Research uses measurable data to formulate facts and uncover patterns in research Huck (2004). So, for quantitative data you will need to decide in what format to present your findings i.e. bar charts, pie charts, histograms etc. You will need to label each table and figure accurately and include a list of tables and a list of figures with corresponding page numbers in your Contents page or Appendices.

Following is a list of characteristics and advantages of using quantitative methods: The data collected is numeric, allowing for collection of data from a large sample size, Statistical analysis allows for greater objectivity when reviewing results and therefore, results are independent of the researcher, Numerical results can be displayed in graphs, charts, tables and other formats that allow for better interpretation, Data analysis is less time-consuming and can often be done using statistical software, Results can be generalized if the data are based on random samples and the sample size was sufficient, Data collection methods can be relatively quick, depending on the type of data being collected, and Numerical quantitative data may be viewed as more credible and reliable, especially to policy makers, decision makers, and administrators (Neuman, & Robson, 2004).

For qualitative data you may want to include quotes from interviews. Any sample questionnaires or transcripts can be included in your Appendices. Qualitative analysis and discussion will often demand a higher level of writing / authoring skill to clearly present the emergent themes from the research. It is easy to become lost in a detailed presentation of the narrative and lose sight of the need to give priority to the broader themes. Creswell (2013) stated that Quantitative research deals in numbers, logic, and an objective stance. Quantitative research focuses on numeric and unchanging data and detailed, convergent reasoning rather than divergent reasoning [i.e., the generation of a variety of ideas about a research problem in a spontaneous, free-flowing manner]. So, quantitative data results are presented in the same way.

Singh, (2007) stated that quantitative data result has its main characteristics: The data is usually gathered using structured research instruments, the results are based on

larger sample sizes that are representative of the population, The research study can usually be replicated or repeated, given its high reliability, Researcher has a clearly defined research question to which objective answers are sought, All aspects of the study are carefully designed before data is collected, Data are in the form of numbers and statistics, often arranged in tables, charts, figures, or other non-textual forms, Project can be used to generalize concepts more widely, predict future results, or investigate causal relationships, Researcher uses tools, such as questionnaires or computer software, to collect numerical data.

In quantitative data result presentation, Bryman and Cramer (2005) explained the following things to keep in mind when reporting the results of a study using quantitative methods: Explain the data collected and their statistical treatment as well as all relevant results in relation to the research problem you are investigating. Interpretation of results is not appropriate in this section, Report unanticipated events that occurred during your data collection. Explain how the actual analysis differs from the planned analysis. Explain your handling of missing data and why any missing data does not undermine the validity of your analysis, Explain the techniques you used to "clean" your data set, choose a minimally sufficient statistical procedure; provide a rationale for its use and a reference for it. Specify any computer programs used, Describe the assumptions for each procedure and the steps you took to ensure that they were not violated, When using inferential statistics, provide the descriptive statistics, confidence intervals, and sample sizes for each variable as well as the value of the test statistic, its direction, the degrees of freedom, and the significance level [report the actual p value], Avoid inferring causality, particularly in nonrandomized designs or without further experimentation, Use tables to provide exact values; use figures to convey global effects. Keep figures small in size; include graphic representations of confidence intervals whenever possible, Always tell the reader what to look for in tables and figures.

Generally, Quantitative methods emphasize objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a particular phenomenon

Summary, Conclusion, And Recommendations

Summary

A research summary is a professional piece of writing

that describes your research to some prospective audience. Main priority of a research summary is to provide the reader with a brief overview of the whole study. To write a quality summary, it is vital to identify the important information in a study, and condense it for the reader. Having a clear knowledge of your topic or subject matter enables you to easily comprehend the contents of your research summary (Philip, 1986).

As Globio (2017), stated that guidelines in writing the summary of findings are the following.

1. There should be brief statement about the main purpose of the study, the population or respondents, the period of the study, method of research used, the research instrument, and the sampling design.

Example, a research conducted study of teaching science in the high schools of Province may be explained as: This study was conducted for the purpose of determining the status of teaching science in the high schools of Province A. The descriptive method of research was utilized and the normative survey technique was used for gathering data. The questionnaire served as the instrument for collecting data. All the teachers handling science and a 20% representative sample of the students were the respondents. The inquiry was conducted during the school year 1989-'90.

2. The findings may be lumped up all together but clarity demands that each specific question under the statement of the problem must be written first to be followed by the findings that would answer it. The specific questions should follow the order they are given under the statement of the problem.

Example. How qualified are the teachers handling science in the high schools of province A?

Of the 59 teachers, 31 or 53.54 % were BSC graduates and three or 5.08% were MA degree holders. The rest, 25 or 42.37%, were non-BSC baccalaureate degree holders with at least 18 education units. Less than half of all the teachers, only 27 or 45.76% were science majors and the majority, 32 or 54.24% were non-science majors.

3. The findings should be textual generalizations, that is, a summary of the important data consisting of text and numbers. Every statement

of fact should consist of words, numbers, or statistical measures woven into a meaningful statement. No deductions, nor inference, nor interpretation should be made otherwise it will only be duplicated in the conclusion.

Only the important findings, the highlights of the data, should be included in the summary, especially those upon which the conclusions should be based. Findings are not explained nor elaborated upon anymore. They should be stated as concisely as possible.

The summary actually is found at the beginning of the written piece and will often lead to a concise abstract of the work which will aid with search engine searches (Erwin,2013). The summary of any written paper that delves into a research related topic will provide the reader with a high level snap shot of the entire written work. The summary will give a brief background of the topic, highlight the research that was done, significant details in the work and finalize the work's results all in one paragraph. Only top level information should be provided in this section and it should make the reader want to read more after they see the summary.

Conclusions

Conclusion is one part of research. Girma Tadess (2014) stated that the conclusion may be the most important part of the research. The writer must not merely repeat the introduction, but explain in expert like detail what has been learned, explained, decided, proven, etc. The writer must reveal the way in which the paper's thesis might have significance in society. It should strive to answer questions that the readers logically, raise. The writer should point out the importance or implication of the research on the area of the societal concern.

Guidelines in writing the conclusions. The following should be the characteristics of the conclusions Philip (1986):

1. Conclusions are inferences, deductions, abstractions, implications, interpretations, general statements, and/or generalizations based upon the findings. Conclusions are the logical and valid outgrowths upon the findings. They should not contain any numeral because numerals generally limit the forceful effect or impact and scope of a generalization. No conclusions should be made that are not based upon the findings.

Example: *The conclusion that can be drawn from the findings in No. 2 under the summary of findings is this: All the teachers were qualified to teach in the high school but the majority of them*

were not qualified to teach science.

2. Conclusions should appropriately answer the specific questions raised at the beginning of the investigation in the order they are given under the statement of the problem. The study becomes almost meaningless if the questions raised are not properly answered by the conclusions.

Example. If the question raised at the beginning of the research is:

“How adequate are the facilities for the teaching of science?” and the findings show that the facilities are less than the needs of the students, the answer and the conclusion should be: “The facilities for the teaching of science are inadequate”.

3. Conclusions should point out what were factually learned from the inquiry. However, no conclusions should be drawn from the implied or indirect effects of the findings.

Example: From the findings that the majority of the teachers were non-science majors and the facilities were less than the needs of the students, what have been factually learned are that the majority of the teachers were not qualified to teach science and the science facilities were inadequate?

It cannot be concluded that science teaching in the high schools of Province A was weak because there are no data telling that the science instruction was weak. The weakness of science teaching is an indirect or implied effect of the non-qualification of the teachers and the inadequacy of the facilities. This is better placed under the summary of implications.

If there is a specific question which runs this way “How strong science instruction in the high schools of Province A as is perceived by the teachers and students?”, then a conclusion to answer this question should be drawn. However, the respondents should have been asked how they perceived the degree of strength of the science instruction whether it is very strong, strong, fairly strong, weak or very weak. The conclusion should be based upon the responses to the question.

4. Conclusions should be formulated concisely, that is, brief and short, yet they convey all the necessary information resulting from the study as required by the specific questions.

Without any strong evidence to the contrary,

conclusions should be stated categorically. They should be worded as if they are 100 percent true and correct. They should not give any hint that the researcher has some doubts about their validity and reliability. The use of qualifiers such as probably, perhaps, may be, and the like should be avoided as much as possible.

5. Conclusions should refer only to the population, area, or subject of the study. Take for instance, the hypothetical teaching of science in the high schools of Province A, all conclusions about the faculty, facilities, methods, problems, etc. refer only to the teaching of science in the high schools of Province A.

Conclusions should not be repetitions of any statements anywhere in the thesis. They may be recapitulations if necessary but they should be worded differently and they should convey the same information as the statements recapitulated.

In drawing the conclusion, we should be aware of Some Dangers to Avoid in Drawing up Conclusions (Bacani, et.al., pp. 48-52) avoid Bias, Incorrect generalization (An incorrect generalization is made when there is a limited body of information or when the sample is not representative of the population), Incorrect deduction (This happens when a general rule is applied to a specific case), Incorrect comparison (A basic error in statistical work is to compare two things that are not really comparable), Abuse of correlation data(A correlation study may show a high degree of association between two variables), Limited information furnished by any one ratio, and Misleading impression concerning magnitude of base variable. So, Conclusions should not be repetitions of any statements anywhere in the thesis. They may be recapitulations if necessary but they should be worded differently and they should convey the same information as the statements recapitulated.

The conclusion will be towards the end of the work and will be the logical closure to all the work found at the end of the document. As Erwin (2013) the conclusion will have more detailed information than the summary but it should not be a repeat of the entire body of the work. The conclusion should revisit the main points of the research and the results of the investigation. This section should be where all the research is pulled together and all open topics be closed. The final results and a call to action should be included in this phase of the writing.

RECOMMENDATIONS

In recommendation section, it should be included in the a report part when the results and conclusions indicate that further work must be done or when the writer needs to discuss several possible options to best remedy

a problem. The writer should not introduce new ideas in the recommendations section, but rely on the evidence presented in the result and conclusions sections. Via the recommendations section, the writer is able to demonstrate that he or she fully understands the importance and implication of his or her research by suggesting ways in which it may be further developed (Berk , Hart , Boerema ,and Hands, 1998).

Furthermore, Erwin, (2013) described that for recommending similar researches to be conducted, the recommendation should be: It is recommended that similar researches should be conducted in other places. Other provinces should also make inquiries into the status of the teaching of science in their own high schools so that if similar problems and deficiencies are found, concerted efforts may be exerted to improve science teaching in all high schools in the country.

Research Report Writing

Research report is a condensed form or a brief description of the research work done by the researcher. It involves several steps to present the report in the form of thesis or dissertation. A research paper can be used for exploring and identifying scientific, technical and social issues. If it's your first time writing a research paper, it may seem daunting, but with good organization and focus of mind, you can make the process easier on yourself (Berk, Hart, Boerema , and Hands, 1998). In addition to this Erwin,(2013) stated that writing a research paper involves four main stages: choosing a topic, researching your topic, making an outline, and doing the actual writing. The paper won't write itself, but by planning and preparing well, the writing practically falls into place. Also, try to avoid plagiarism. So, in each section of the paper we will need to be critically writing the paper. Most of the consideration in each section explained as follow:

a. Introduction: The introduction is a critical part of your paper because it introduces the reasons behind your paper's existence. It must state the objectives and scope of your work, present what problem or question you address, and describe why this is an interesting or important challenge(Erwin, 2013). In addition, it is important to introduce appropriate and sufficient references to prior works so that readers can understand the context and background of the research and the specific reason for your research work. Having explored those, the objectives and scope of your work must be clearly stated. The introduction may explain the approach that is characteristic to your work, and mention the essence of the conclusion of the paper.

b. Methods: The Methods section provides

sufficient detail of theoretical and experimental methods and materials used in your research work so that any reader would be able to repeat your research work and reproduce the results. Be precise, complete and concise: include only relevant information. For example, provide a reference for a particular technique instead of describing all the details.

c. Results: The Results section presents the facts, findings of the study, by effectively using figures and tables. Wilkinson (1991) explained that this section must present the results clearly and logically to highlight potential implications. Combine the use of text, tables, and figures to digest and condense the data, and highlight important trends and extract relationships among different data items. Figures must be well designed, clear, and easy to read. Figure captions should be succinct yet provide sufficient information to understand the figures without reference to the text.

d. Discussion: In the Discussion section, present your interpretation and conclusions gained from your findings. You can discuss how your findings compare with other experimental observations or theoretical expectations. Refer to your characteristic results described in the Results section to support your discussion, since your interpretation and conclusion must be based on evidence. By properly structuring this discussion, you can show how your results can solve the current problems and how they relate to the research objectives that you have described in the Introduction section. This is your chance to clearly demonstrate the novelty and importance of your research work (Wilkinson, 1991).

e. Conclusions: The Conclusion section summarizes the important results and impact of the research work. Future work plans may be included if they are beneficial to readers (Singh, 2007).

f. Acknowledgments: The Acknowledgments section is to recognize financial support from funding bodies and scientific and technical contributions that you have received during your research work.

g. References: The References section lists prior works referred to in the other sections. It is vitally important from an ethical viewpoint, to fully acknowledge all previously published works that are relevant to your research. Whenever we use previous knowledge, we must acknowledge the source. Readers benefit from complete references as it enables them to position our work in the context of current research. Ensure that the references given are sufficient as well as current, and

accessible by the readers (Singh, 2007).

Writing a research paper need critical attention; in the area different expertise stated that research report writing need critical rewriting, editing, revising and etc. stages to make it more effective, accurate and acceptable. Among them Philip (1986) stated the following tips may be useful in writing the paper:

You need not start writing the text from the Introduction. Many authors actually choose to begin with the results section since all the materials that must be described are available. This may provide good motivation for carrying out the procedure most effectively. And Your paper must be interesting and relevant to your readers. Consider what your readers want to know rather than what you want to write. Describe your new ideas precisely in an early part of your paper so that your results are readily understood. Otherwise, do not use lengthy descriptions of the details. For example, writing too many equations and showing resembling figures or too much detailed tables should be avoided. Clarity and conciseness are extremely important.

He also added that during and after writing your draft, you must edit your writing by reconsidering your starting plan or original outline. You may decide to rewrite portions of your paper to improve logical sequence, clarity, and conciseness. This process may have to be repeated over and over. When editing is completed, you can send the paper to your co-authors for improvement. When all the co-authors agree on your draft, it is ready to be submitted to the journal. It is worth performing one final check of grammatical and typographical errors. English correction of the manuscript by a native speaker is highly recommended before your submission if you are not a native speaker. Unclear description prohibits constructive feedback in the review process.

In relation to Writing and Editing Philip (1986);Singh (2007); Wilkinson(1991) and (Erwin, 2013). and the following tips may be useful in writing the paper.

The first, we need not start writing the text from the Introduction. Many authors actually choose to begin with the results section since all the materials that must be described are available. This may provide good motivation for carrying out the procedure most effectively.

The Second, Our paper must be interesting and relevant to our readers. Consider what our readers want to know rather than what we want to write. Describe our new ideas precisely in an early part of our paper so that our results are readily understood. Otherwise, we do not use lengthy descriptions of the details. For example, writing too many equations and showing resembling figures or too much detailed tables should be avoided. Clarity and conciseness are extremely important.

The third, during and after writing our draft, we must

edit our writing by reconsidering our starting plan or original outline. We may decide to rewrite portions of our paper to improve logical sequence, clarity, and conciseness. This process may have to be repeated over and over.

The fourth, when editing is completed, we can send the paper to our co-authors for improvement. When all the co-authors agree on the draft, it is ready to be submitted to the journal (if journal publication is needed).

It is worth performing one final check of grammatical and typographical errors.

The fifth, English correction of the manuscript by a native speaker is highly recommended before our submission if we are not a native speaker. Unclear description prohibits constructive feedback in the review process.

Research report has its own format or organization which is accepted by in different field of study John (1970) stated that research has the following Format:

Preliminary Section

This part includes: Title Page (Be specific. Tell what, when, where, etc. In one main title and a subtitle, give a clear idea of what the paper investigated), Acknowledgments (if any) (Include only if special help was received from an individual or group), Table of Contents (Summarizes the report including the hypotheses, procedures, and major findings), List of Tables (if any), List of Figures (if any) and Abstract.

Main Body

CHAPTER ONE: INTRODUCTION

This part of the paper includes: Background of the study (overview of the study and This is a general introduction to the topic), Statement of the problem (This is a short reiteration of the problem), Objectives of the study (What is the goal to be gained from a better understanding of this question?), Scope of the study, Limitation of the study (Explain the limitations that may invalidate the study or make it less than accurate), Significance of the study (Comment on why this question merits investigation), Origination of the study, and Definition of Terms (Define or clarify any term or concept that is used in the study in a non-traditional manner or in only one of many interpretations).

CHAPTER TWO: REVIEW OF RELATED LITERATURE

This part of the thesis states analysis of previous research and It Gives the reader the necessary background to understand the study by citing the investigations and findings of previous researchers and

documents the researcher's knowledge and preparation to investigate the problem.

CHAPTER THREE: METHODOLOGY

The methodology part includes: Design of the study (Description of Research Design and Procedures Used), Sample method and size/ Sampling Procedures, Sources of Data (Give complete information about who, what, when, where, and how the data were collected), Methods and Instruments of Data Gathering (Explain how the data were limited to the amount which was gathered. If all of the available data were not utilized, how was a representative sample) achieved? , and Gives the reader the information necessary to exactly replicate (repeat) the study with new data or if the same raw data were available, the reader should be able to duplicate the results. This is written in past tense but without reference to or inclusion of the results determined from the analysis.

CHAPTER FOUR: DATA ANALYSIS

It contains: text with appropriate, tables and figures. Describe the patterns observed in the data. Use tables and figures to help clarify the material when possible.

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION

Under this part of our paper we will include: Restatement of the Problem , Description of Procedures, Major Findings (reject or fail to reject H_0) , Conclusions, Recommendations for Further Investigation, and This section condenses the previous sections, succinctly presents the results concerning the hypotheses, and suggests what else can be done.

A. Reference Section: includes: End Notes (if in that format of citation), Bibliography or Literature Cited and Appendix.

SUMMARY

In summing up, research is a scientific field which helps to generate new knowledge and solve the existing problem. So, to get this function we need to pass deferent stages. Among this data analysis is the crucial part of research which makes the result of the study more effective. Data Analysis is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information. The results so obtained are communicated, suggesting conclusions, and supporting decision-making. Data analysis is to verify our results whether it is valid, reproducible and

unquestionable and it is a process used to transform, remodel and revise certain information (data) with a view to reach to a certain conclusion for a given situation or problem.

It can be applied in two ways which is qualitatively and quantitative. Whatever the research apply one of the two in his or her research applying the most effective data analysis in research work is essential. Due to this, data analysis is beneficial because it helps in structuring the findings from different sources of data collection like survey research, is again very helpful in breaking a macro problem into micro parts, and acts like a filter when it comes to acquiring meaningful insights out of huge data-set. Furthermore, every researcher has sort out huge pile of data that he/she has collected, before reaching to a conclusion of the research question. Mere data collection is of no use to the researcher. Data analysis proves to be crucial in this process, provides a meaningful base to critical decisions, and helps to create a complete dissertation proposal. So, after analyzing the data the result will provide by qualitative and quantitative method of data results.

In research work, summary, conclusion and recommendation are the most important part which is need to be write in effective and efficient to make the paper more convincible and reputable. And in writing research report needs a critical attention to make the report more academic and effective.

Generally, one of the most important uses of data analysis is that it helps in keeping human bias away from research conclusion with the help of proper statistical treatment. With the help of data analysis a researcher can filter both qualitative and quantitative data for an assignment writing projects. Thus, it can be said that data analysis is of utmost importance for both the research and the researcher.

REFERENCES

- Ackoff, R.L.(1961).*The Design of Social Research*. Chicago: University of Chicago Press.
- Addison,W. (2017). *Medical Statistics Course*. MD/PhD students, Faculty of Medicine.
- Agresti, A. & Finlay, B.(1997). *Statistical Methods for the Social Sciences (3rd ed.)*.Prentice Hall.
- Berk, M., Hart,B., Boerema,D., and Hands, D. (1998). *Writing Reports: Resource Materials for Engineering Students*.University of South Australia.
- Bryman,A and Cramer,D. (2005). *Quantitative Data Analysis with SPSS 12 and 13: A Guide for Social Scientists*. London: Routledge.
- Bussines Dictionary .(2017). *Data Analysis*. Retrieved from <http://www.businessdictionary.com/definition/data-analysis.html>
- Celine.(2017)."*Difference between Population and Sample*." Retrieved on May, 5/2018.from<http://www.differencebetween.net/miscellaneous/difference-between-population-and-sample/>
- Cohen , L., Manion, L. and Morrison, K. (2007). *Research Methods in Education*. London and New York: Routledge.
- Cohen, J.W. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nded.)*. London and New York: Routledge.
- Creswell, J. W. (2002). *Educational Research: Planning, Conducting, and Evaluating Quantitative*. Prentice Hall.
- Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed methods Approaches*. Sage Publications, Incorporated.
- Daniel,M.(2010). *Doing Quantitative Research in Education with SPSS (2nded.)*. London: SAGE Publications.
- Dawson, C.(2002). *Practical Research Methods: A user-friendly Guide to Mastering Research Techniques and Projects*. Oxford: How to Books Ltd,
- Denscombe , M. 2007. *The Good Research Guide for Small-scale Social Research Projects (3rded.)*.Open University Press: McGraw-Hill Education
- Earl,R.B.(2010). *The Practice of Social Research (12thed.)*. Belmont, CA: Wadsworth Cengage.
- Erwin,M. G.(2013) . *Thesis Writing: Summary, Conclusions, and Recommendations*. Retrieved from <http://thesisadviser.blogspot.com/2013/02/thesis-writing-summary-conclusions-and.html>
- Field, A. (2002). *Discovering Statistics Using SPSS for Windows*. London: Sage
- Freeman,J and Young,T.(2017), *Correlation Coefficient: Association between Two Continuous Variables*. Retrieved on May, 2017 from <http://www.epa.gov/bioindicators/statprimer/index.html>
- Gogtay,N.J., Deshpande,S., and Thatte,U.M.(2017). *Principles of Correlation Analysis*. J AssocPhyInd 2017; 65:78-81.
- Hill,M.H.(2013). *Format of Research Reports*.adapted from: John W. Best, Research in Education, 2nd ed., (Englewood Cliffs, NJ: Prentice-Hall, 1970)]. Retrieved from <http://www.jsu.edu/dept/geography/mhill/research/researchf.html>
- Hinkle, D. E., Wiersma, W., &Jurs, S. G. (2003).*Applied Statistics for the Behavioral Sciences (5th ed.)*. Boston, MA: Houghton Mifflin Company.
- Howell, D. C. (2002).*Statistical Methods for Psychology (5thed.)*. Pacific Grove, CA: Duxbury. Retrieved from https://www.sheffield.ac.uk/polopoly_fs/1.43991!/file/Tutorial-14-correlation.pdf.
- Huck, S. W. (2004).*Reading Statistics and Research (4thed.)*. Boston, MA: Allyn and Bacon.
- Juliet,C.J and Nicholas,L. H.(2005). *Ground Theory*. London Thousand Oaks: New Delhi.

- Kothari, C.R. (2004). *Research methodology (2nd ed)*. New Delhi: New Age international (p) limited publisher.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for Intermediate Statistics: Use MED819, ANCOVA*. Retrieved from <http://www.mas.ncl.ac.uk/~njnsm/medfac/doc/ancova.pdf> on May, 5/2018.
- Mordkoff, J.T. (2016). *Introduction to ANOVA*. Retrieved from <http://www2.psychology.uiowa.edu/faculty/mordkoff/GradStats/part%203a/Intro%20to%20ANOVA.pdf>
- Nathaniel, E. H. (2017). *Analysis of Covariance*. Retrieved from <http://users.stat.umn.edu/~helwig/notes/acov-Notes.pdf>
- Neuman, W.L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches (4th ed.)*. Boston: Allyn and Bacon.
- Neuman, W. L., & Robson, K. (2004). *Basics of Social Research*. Pearson.
- Norgaard, R. (n.d). *Results and Discussion Sections in the Scientific Research Article*. University of Colorado at Boulder. Retrieved on May, 10, 2018 resess.unavco.org/lib/downloads/RESESS.13.results&discussion.051313.pdf
- O'Brien, D. and Scott, P.S. (2012). "Correlation and Regression", in *Approaches to Quantitative Research – A Guide for Dissertation Students*, Ed, Chen, H, Oak Tree Press.
- Ott, R.L. 1993. *An Introduction to Statistical Methods and Data Analysis (4th ed.)*. Belmont, CA: Duxbury Press,
- Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods (2nd ed.)*. Newbury Park, CA: Sage.
- Philip, C.K. (1986). *Successful Writing at Work (2nd ed.)*. Retrieved from [ftp://nozdr.ru/biblio/kolxo3/L/LEn/Kolin%20P.%20Successful%20writing%20at%20work%20\(Wadsworth,%202009\)\(ISBN%200547147910\)\(O\)\(753s\)_LEn_.pdf](ftp://nozdr.ru/biblio/kolxo3/L/LEn/Kolin%20P.%20Successful%20writing%20at%20work%20(Wadsworth,%202009)(ISBN%200547147910)(O)(753s)_LEn_.pdf)
- Quirk T.J., and Rhiney, E. (2016). *Multiple Correlations and Multiple Regressions*. Excel for Statistics. Springer, Cham.
- Selltiz, C., Jahoda, M., Deutsch, M., and Cook, S.W. (1959). *Research Methods in Social Relations (rev. ed.)*. New York: Holt, Rinehart and Winston, Inc.
- Singh, K. (2007). *Quantitative Social Research Methods*. Los Angeles, CA: Sage Publications.
- Strauss, A and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.
- The Advanced Learner's Dictionary of Current English, **Oxford, 1952, p. 1069.**
- Weiss, N.A. (1999). *Introductory Statistics*. Addison Wesley.
- Wilkinson, A. (1991). *The Scientist's Handbook for Writing Papers and Dissertations*. Englewood Cliffs, New Jersey: Prentice Hall.