GUIDE IN STATISTICS FOR TERTIARY INSTITUTIONS RESEARCH

NA'IYA K. H[1]; AMINU Z. G[1]; RABIU S. R[2]; NALADO T. K[3]; & ABUBAKAR B. T[4]

[1]Department of Geography, Federal College of Education Katsina
[2]Department of computer science, Federal College of Education Katsina [3] Director ICT & Vocational Services, Katsina State Ministry of Education
[4]Asst. Zonal Coordinator & Head of Operations, zone 2, katsina. National Board for Arabic and Islamic Studies.

Abstract

Statistics is extensively used in every discipline for meaningful conclusions from the data. This work discusses the importance of learning statistics to students and other researchers in tertiary institutions. The work tries to give the teachers and students a broader picture as to how different types of descriptive or inferential statistical processes can be used to address different types of research questions in tertiary institutions. Statistical skills and other methods for analyzing data were use with simple examples to arrive at the solutions required test. Statistics methods do vary and they depend on the immediate requirement amongst the researchers and type of research questions generated and answers expected. This work contains all the material that would be taught in any statistics course meant for research undertaking. Methods and statistical tools for comparism, t-test, correlational two sample test, chi square, and ANOVA. Sources and methods of data collection process were discussed. By the end of the work, teachers/students will have the understanding needed to succeed in statistics, and if one is studying independently then by the end of this work, one will have a well-rounded coherent, understanding of statistics and probability in statistical course. The paper starts with the absolute basics so that, anyone at any level can comprehend the work with basics like; mean, mode medium, sampling methods, averages and standard deviation and then we will get into probability theory including, conditional probability, leading to conclusion of a given sample data.

Keyword: *Statistics, descriptive and inferential statistics, t-test, methods and chi-square*

INTRODUCTION

Meaning of Statistics

There are different definitions of Statistics; statistics is concerned with scientific methods of collecting, organizing, summarizing, presenting and analyzing data as well as drawing valid conclusion on the basis of such analysis. Statistics is a very broad subject, with applications in a vast number of different fields. Statistical analysis uses summary numbers by organizing them, observing them, and od inferring their relationships to each other. Almost any summary statement concerning the information in a database is likely to contain a statistic. (Adeyemi, 2002)

Concept of Statistics

Generally, one can say that statistics is the methods for collecting, analyzing, interpreting and drawing conclusions from information. Statistics is the science of knowledge that deals with the collection, analyzing, summarizing, presenting, interpreting and the use of data in order to aid decision making study (Baddie and Halley, 1995; Kolawole, 2001 in Adeyemi, 2009). It is the methods which scientists and mathematicians have developed for interpreting and drawing conclusions from collected data

Generally, Statistics serves in two capacities:

i. It gives methods for organizing, summarizing and communicating data and
ii. It provides methods for making inference beyond the observation.

In summary therefore, Statistics involves observation, collection of data, organization of data, presentation of data, interpretation of data for decision making. It should be noted that Statistics as a subject is not the plural of Statistic. Statistic is a measure which we obtain by observing and studying the characteristics of the sample.

The raw material of statistics is data. For this purpose, data can be defined as numbers. The two kinds of numbers that we use in statistics are numbers that result from taking a measurement and those that result from the process of counting. For example, when a nurse weighs a patient or takes a patient's

temperature, a measurement, consisting of a number such as 30 kg or 37^0 C, is obtained. A different type of number is obtained when a hospital administrator counts the number of patients perhaps 15 discharged from the hospital on a given day. Each of these three numbers is a datum, and the three numbers taken together are data. (Adeyemi, 2009).

Types of Statistics

Statistics can be classified into two main branches.

1. Descriptive Statistics: Descriptive Statistics is concerned with the organization and presentation of data in a convenient, usable and communicable form. It is the set of methods serving the function s of organizing, summarizing and communicating data. Some of these methods include measures of central tendency, measures of variations.
2. Inferential Statistics: In this method's inferential Statistics, the aim is to make a decision about a population based on a sample from the population. These involve statistical methods used for arriving at conclusions extending beyond immediate data. There are two main methods used in inferential statistics: estimation and hypothesis testing. (Kothari, 2004).

Benefits/Importance of the Study of Statistics

When you study statistics, you will acquire useful knowledge, skills, capabilities or dispositions that vary according to the extent and level of study, or training in the subject. Some of these benefits according to Songsore, (2018) include that the study of statistics will enable the learner to:

1. Acquire knowledge and skills in observation, collection, organization, communication, analysis of data, drawing inferences from the analysis and making sound decisions;
2. Read, understand and interpret communicated data, follow inferences drawn therefrom and appreciate decisions made consequent upon the inference drawn;
3. Successfully execute empirical research. Empirical research can be carried out for answering research questions, testing hypotheses or taking decisions and making predictions;

4. Read, interpret and make use of research reports or articles for decision making;
5. Acquire the skills and techniques for estimating, predicting and projecting into the future on the previous and present data; and
6. Draw sound conclusions based on the data or information obtained.

Types of Statistics

There are two major types of statistics. The branch of statistics devoted to the summarization and description of data is called descriptive statistics and the branch of statistics concerned with using sample data to make an inference about a population of data is called inferential statistics.

1. Descriptive Statistics: Descriptive statistics consist of methods for organizing and summarizing information (Weiss, 1999). 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.
2. Inferential Statistics: 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 includes methods like point estimation, interval estimation and hypothesis testing which are all based on probability theory.

Differences between Descriptive and Inferential Statistics

Although descriptive and inferential Statistics both are used for the purpose of analysis of data, still both of them are used differently in various ways.

1) Descriptive Statistics gives a description about a sample, while the inferential statistics predicts and infers about a much larger data or population.
2) Descriptive Statistics just describes the certain characteristics about a data whereas inferential Statistics deeply analyze the statistical data and observation.
3) Descriptive Statistics deals with central tendency and spread of the distribution. While in inferential Statistics, more details such as hypothesis tests and confidence interval are studied.

4) The measures of descriptive Statistics (mean, median and mode) are numbers on the other hand, the measures in inferential Statistics are not always exact numbers.
   In this methods Statistics, conclusions cannot be made beyond the given data. In inferential Statistics, the educated predictions and guesses can be made on the basis of the parameters of a given population; it does not matter how big the sampled population is.

Tests for Inferential Statistics

T–Test: Can be used as an inferential method to compare the mean of the sample to the population mean using z-scores and the normal probability curve. One uses t-curves for various degrees of freedom associated with one's data. Degrees of freedom are the number of observations that vary around a constant. You can use t-test to compare two means.

ANOVA: Analysis of variance is a ratio of observed differences between more than two means. –It is more versatile than a t-test and should be used in most cases in lieu of the T-test. –The analysis allows comparison of means of the samples and testing of the null hypothesis regarding no significant difference between means of the samples.

Chi-square: An index used to find the significance of differences between the proportions of subjects, events, objects that can be stratified into different categories.

Procedure for Carrying Out Inferential Statistics

Some step-by-step procedure for performing inferential statistics.

1) Start with a theory
2) Make a research hypothesis
3) Operationalize the variables (define and measure a specific variable as it is used in your study).
4) Identify the population to which the study results should apply
5) Form a null hypothesis for this population
6) Collect a sample from the population and run the study
7) Perform statistical tests to see if they obtained sample characteristics are sufficiently different from what would be expected under the null hypothesis to be able to reject the null hypothesis. (Queiroz, 2017)

Problems Associated with Inferential Statistics

1. Problem of the p value.
   The p value is at center of most applications of inferential statistics. Its major problem may be that it is not intuitive. Few investigators seem to know what it means when it is low; even fewer know what it means when it is high.
2. Problem of making conclusions
   In inferential statistics, conclusions often contain the risk of bias. When investigators make conclusions after an attempt to prove a hypothesis, they have acted with bias. Such conclusions need not be trusted readily since a lot of educated guesses are often involved.
3. Problem of power estimates
   Power estimates are required in some settings. If you plan to use inferential statistics to analyze your evaluation results, you should first conduct a power analysis to determine what size sample you will need.

Variables

A variable on inferential Statistics is any characteristic that varies from one individual member of the population to another. In other words, any type of observation which can take different values for different people, or different values at different times, or places, is called a variable. The following are examples of variables in sample:

1. Family size, number of hospital beds, date of birth, number of schools in a country, etc.
2. Height, mass, blood pressure, temperature, blood glucose level, etc. according to (Songsore, 2018)

Types of Variables

There are two types of variables.

1. Quantitative variables: A quantitative variable is one that can take numerical values. The variables (Family size, number of hospital beds, date of birth, number of schools in a country, height, mass, blood pressure, temperature, blood glucose level, etc) are examples of quantitative variables. Quantitative variables may be characterized further as to whether they are discrete or continuous.

i. Discrete variables: A variable is discrete if it can assume only a infinite number of values or as many values as there are integers (whole numbers). The variables in (1), above, can be counted. These are examples of discrete variables. A discrete variable is characterized by gaps or interruptions in the values that it can assume. The following example illustrates the point. The number of daily admissions to a hospital is a discrete variable since it must be represented by a whole number, such as 0, 1, 2 or 3. The number of daily admissions on a given day cannot be a number such as 1.8, 3.96 or 5.33.

ii. Continuous variables: The variables in (2), above, can be measured. These are examples of continuous variables. A continuous variable does not possess the gaps or interruptions characteristic of a discrete variable. A continuous variable can assume any value within a specific relevant interval of values assumed by the variable

2. Qualitative variables: Variables which cannot take numerical values are called qualitative variables. A qualitative variable can neither be measured nor be counted. The following are examples of qualitative variables: place of birth, nationality, color, color of hair, gender, blood group, smoking habit, surname, rank in military. (Verma 2019)

SOURCES OF STATISTICAL DATA

Adeyemi, (2002) postulate that, the raw material of statistics is data. Hence, to get data in statistics, two main sources are identified. These are: primary sources and secondary sources.

Primary Sources of Data

Here there are two broad sources – experiments and surveys.

1. Experiments: Frequently, the data needed to answer a question are available only as a result of an experiment. A researcher may wish to know which of several drugs is most effective for treating headache. The researcher might conduct an experiment by assigning the drugs to different patients. Subsequent evaluation of the responses to the different

drugs might enable the researcher to decide which drug is most effective for treating headache.

2. Surveys: In surveys, the aim of the researcher is to find a way of obtaining information from individuals, referred to as respondents. Such information can be factual (for example, the number of cars per household, age of respondents, or income) or can concern the attitudes of the respondent, for example, his attitude to racial discrimination, or his liking for a brand of cigarette. (Ololobou,2007)

Advantages

- ❖ It makes wide geographic coverage possible at comparatively little cost.
- ❖ There is no need to train interviewers.
- ❖ It encourages the respondent to answer questions frankly in the privacy of the home and without the subjective influence of the interviewers.
- ❖ There is lack of interviewer bias.

Disadvantages

- ❖ One cannot be sure of the interpretation placed by the respondent on the questions asked.
- ❖ There may be a delay in receiving responses.
- ❖ There is the problem of non-response to the survey. This non-response is certain to affect the validity of the survey as it is most unlikely that the sections of the sample that do and do not reply are similar in the characteristics under consideration.

Secondary Sources of Data

Secondary data are data originally not collected under the supervision of the person or organization using the data. Secondary data are available from libraries, government agencies and the internet. (Kolawole, 2001)

Library: A common place to look for secondary data is a library. Here, data can be obtained from magazines, text books, journals, encyclopedias and newspapers.

Government agencies: Government data can be obtained from publications issued by local, state, national and international

governments. Such data include laws, regulations, Statistics and consumer information.

Internet: Secondary data can be obtained from search engines such as Yahoo, Google, MSN.com, etc., on the internet.

Advantages of Secondary Data

- Immediately available.
- Cheaper than obtaining new data.

Disadvantages of Secondary Data

- May be incomplete.
- May have been collected to satisfy different needs.
- No control exists over the method of collection and accuracy of the data.

Methods of Data Collection

a. Questionnaire

Questionnaire is a set of questions that have been prepared to ask a number of questions and collect answers from respondents relating to a research topic Silverman, (2000). Respondents are expected to write their answers next to the questions. Properly constructed and responsibly administered questionnaires can become a vital instrument by which the view of a few group of people can represent the opinion of an entire population. The purposes of questionnaire are:

- to collect appropriate data relevant to a research
- to make data comparable and amenable to analysis
- and to minimize bias in formulating and asking questions.

b. Interviews

There are three types of interview; Structured, Semi-structured and Unstructured. According to Gill, Stewart, Treasure and Chadwick, (2008), Structured interviews are in most cases verbally administered questionnaires where a list of predetermined questions are asked, with little or no variation and with no scope for follow-up questions or responses that will require further explanation. Consequently, they are relatively quick and easy to administer.

While unstructured interviews do not reflect any preconceived theories or ideas or follow any pattern. This type of interview may simply start with an opening question such as 'Can you tell me about your experience as a Social Studies student?' and will then progress based, primarily, upon the initial response. Unstructured interviews are usually very time-consuming and can be difficult to manage because of the lack of predetermined interview questions which could have provided little guidance on what to talk about.

Semi-Structured interviews consist of Structured questions that can help to define the areas to be explored, but also allows the interviewer or interviewee to digress in order to pursue an idea or response in more detail.

The following procedures could be observed while conducting an interview-

i. The purpose of the research should be explained to the respondents.
ii. Respondents should be given assurance about ethical principles such as anonymity and confidentiality.
iii. Interviews should be conducted in areas free from distractions and at times locations that are most suitable for participants.
iv. Establishing rapport with respondents before the interview can make it hitch-free.
v. At the end of the interview, the interviewer should thank the respondents for their time and cooperation and ask them if there was anything they would like to add.

c. Observation method

This is another method of data collection where the researcher has to put himself in the actual situation and watch the test subjects carefully. According to Attride-Stirling, (2001), the success of the observation entirely depends on the talents of the researcher since he has to rely on his ability to be vigilant, his intuition, knowledge, skills and experience to collect the data without interacting with the respondents. Observation method often requires long period of time, (months or years) of intensive work because the researcher needs to become accepted as a natural part of the subjects to be sure that their behaviors are natural.

d. Document review and analysis

Document review is a way of collecting data by reviewing existing documents. O'Leary, (2014) classified documents that can be reviewed into-

i. Public records: these are official, ongoing records of an organization's activities. Examples are past results of students, mission statements, policy manuals, student handbooks, strategic plans and so on.
ii. Personal documents: these are first – person accounts of an individual's actions, experiences and beliefs. They include calendars, dairies, e-mails, scrapbooks, duty logs, reports, journals and so on.
iii. Physical evidence: they are physical objects found within the study setting, like flyers, posters, agendas, training materials and so on. (Ali, 2011)

e. Focus groups

A focus group is a group interview of approximately six to twelve people who may share similar characteristics or common interests or may be as diverse as having different ages, occupations or belonging to different social class (Basu, 2004). The success of a focus group depends on how the group members interact a[mong themselves. Thus, the researcher has to create an environment that will encourage the participants to share their perceptions and points of view and guide them to respond effectively to predetermined set of topics.

f. Survey Method

This method is used in different areas for collecting data especially in public and private sectors. In a survey method, the researcher can contact his respondents personally, through telephone or mail (Salant and Dillman,1994). This method takes a lot of time and money but the data collected are of high accuracy, current and in most cases, relevant to the topic. Survey method is flexible, and can be an efficient way of collecting information from a large number of respondents. They are standardized and so are relatively free from several types of errors.

DESCRIPTIVE DATA ANALYSIS

Descriptive Statistics involves the collection, analyzing and informative presentation of a mass of numerical data. Descriptive Statistics help us to simplify large amounts of data in a sensible way.

Other functions of Descriptive Statistics according to (Verma, 2019) include:

1. Providing a powerful summary that may enable comparisons across people or other units.
2. Describing the event or outcome without drawing any conclusion.

Statisticians regard descriptive statistics as data analysis without probability. Attributes such as the mean, mode, median, geometric mean, harmonic mean, range, mean deviation, standard deviation, percentile, kurtosis, proportion and correlation co-efficient etc. fall within the domain of descriptive statistics. Basically, descriptive statistical methods can be divided into main groups: Measures of Central Tendencies and Measures of Dispersion.

Measures of Central Tendency

This is a measure of location, it provides Statistical information about the average, middle and center of a set of data. It is normally a single value that attempts to describe a set of data by identifying the central position within a set of data. There are also classed as summary Statistics.
Three of such measures are the Mean, the Mode, and the Median.

The Mean

This is the arithmetic average of a set of data usually applicable to quantitative data. To obtain the Mean, we sum up all the scores in a set of data to be divided by the number of scores. In other words, mean is the sum of the values, divided by the total number of values. This is given by the simple formula below.

$$\bar{x} = \frac{\sum x}{n}$$

Where, $\bar{x}$ represent the, x is any data value from the data set, $\sum x$ is the sum of all observations and *n* is the total number of data (*n* is called the sample size)

Example 1: The marks obtained by 8 students in an examination are: 27, 32, 33, 45, 59, 63, 75 and 90. The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{27+32+33+45+59+63+75+90}{8} = \frac{424}{8} = 53$$

Hence the mean (Average) mark of the students is 53.

In a situation where the data is in form of frequency table, that is, each data item has frequencies, or if the observations occur frequently then, the following formula is applied.

$\bar{x} = \frac{\sum fx}{\sum f}$, where, $\sum fx$ *is the Sum of* fx and $\sum f$ *is Sum of* f

Example 2: The table below shows the marks obtained by a group of students in a class test.

Marks (x)	0	1	2	3	4	5
Frequency (f)	1	4	9	8	5	3

Calculate the mean mark.

Solution

Marks (x)	0	1	2	3	4	5	
Frequency (f)	1	4	9	8	5	3	30
Fx	0	4	18	24	20	15	81

$\bar{x} = \frac{\sum fx}{\sum f} = \frac{81}{30} = 2.7$ Hence, the mean mark is 2.7

Note: The answer is 2.7; the variable is a continuous variable, so the answer could be in decimal form.

Mean from Grouped Data with Class Intervals

In some cases, the data may be in form of grouped frequency table with class intervals. The following example shows how to calculate the mean for this type of data.

Example 3: In a certain farm, the masses (in KG) of Goats are represented in the following distribution table.

Masses in Kg	20-29	30-39	40-39	50-59	60-69	70-79	80-89
Population	3	4	10	13	11	4	5

The work is arranged in the following table;

Masses (Kg)	Mid-Value (x)	F	Fx
20-29	24.5	3	73.5

30-39	34.5	4	138
40-49	44.5	10	445
50-59	55.5	13	708.5
60-69	65.5	11	709.5
70-79	75.5	4	298
80-89	85.5	5	422.5
		$\sum f = 50$	$\sum fx = 2795$

$$Mean,\ \bar{x} = \frac{\sum fx}{\sum f} = \frac{2795}{50} = 55.9\,kg$$

Hence, the mean mass of the Goats is 55.96kg. Please note that the variable involved is continuous variable, so you can leave the answer in decimal form, unless requested by the questions.

The Median

In a simple term, the median divides the set of data into two equal halves. But before identifying this middle value, the data set must be arranged in order of magnitude. The order could be ascending or descending.

Example 1: The number of children in seven households are: 3, 2, 4, 1, 5, 3 and 6.

From the above, there are seven households. Hence, this is an odd number of items i.e. 7.

Arranging in ascending order of magnitude gives: 1, 2, 3, 3, 4, 5, 6.

Arranging in descending order of magnitude gives: 6, 5, 4, 3, 3, 2, 1.

In each case, the middle number is 3. Hence, the median = 3

Example 2: Find the median of the following set of data: 3, 5, 32, 6, 13, 11, 8, 19, 21, 6.

Arranging in ascending order gives: 3, 5, 6, 6, 8, 11, 13, 19, 21, 32.

Arranging in descending order gives: 32, 21, 19, 13, 11, 8, 6, 6, 5, 3.

Here there set of items, hence no single number at the middle. In this case find the average of the two middle numbers. In each case we have:

$$median = \frac{11+13}{2} = \frac{13+11}{2} = \frac{24}{2} = 12$$ Hence, the median is 12.

The Median can also be calculated from a frequency table. The procedure involves the computation of cumulative frequency, as demonstrated in the following example.

Example 3: The table below shows the marks obtained by a group of students in a class test.

Marks (x)	0	1	2	3	4	5
Frequency (f)	1	4	9	8	5	3

Calculate the median mark

Solution

Marks (x)	Frequency (f)	Cumulative Frequency (C.F.)
0	1	1
1	4	5
2	9	14
3	8	22
4	5	27
5	3	30
	30	

Total frequency, N = 30

The median lies at the $\frac{(30+1)^{th}}{2}\ position$ $\quad i.e.\ \frac{31^{th}}{2} 15.5^{\ th}\ position$

From the cumulative frequency, CF, 15.5^{th} position corresponds to mark 3. Hence, the median is 3.

The Mode

This is the item that occurs most in a set of data or list of items. It is the item or value that has the highest frequency in frequency table. It refers to the most common attributes or value of a variable in which case it is possible for a set of data to have more than one Mode.

Example: Find the mode from the following set of numbers: 2, 2, 3, 2, 3, 4 and 2.

Solution:

The item that occurs most is 2 and hence the mode = 2.

A set of data may have two most frequent values. Bimodal values occur when two or more figures have the same highest frequency. Then, the mode is obtained by taking the arithmetical mean of the two values.

Example: Find the mode of the data below:

2, 3, 3, 4, 2, 6, 3, 3, 4, 4, and 4

Solution: The numbers 3 and 4 have the same highest frequency and therefore, the average of these gives the mode of the numbers.

i.e. $\frac{3+4}{2}=\frac{7}{2}=3.5$

Hence, the mode is 3.5

However, if an ungrouped data is given and that the table has frequencies for the individual values, then the mode is the number that has the highest frequency on the table.

Example: Given the table below, the modal mark

Marks (x)	0	1	2	3	4	5
Frequency (f)	1	4	9	8	5	3

. Solutions the modal mark is 2, since the highest frequency, 9, corresponds to mark 2.

Computation of Mode from Grouped Frequency Data with Class Intervals

Again, mode can be computed from grouped frequency data. In this case the following formula is used.

$$Mode = L_1 + \left(\frac{fx}{fx - fy}\right)c$$

Where, L_1 is the Lower-class boundary of the class containing the Mode;

f_x is the difference between the frequency of the class containing the mode and the frequency before it;

f_y is the difference between the frequency of the class containing the mode and the frequency after it;

c is the class size of the class containing the mode;

Example: Suppose we used the table below, to compute the Mode.

Class Interval	6-10	11-15	16-20	21-25	26-30
Population	6	7	8	3	1

Here, the modal class is 16 – 20, since the highest frequency is 8. Hence, we have;

$L_1 = 15.5,\ f_x = 8-7 = 1,\ f_y = 8-3 = 5,\ c = 5$

Therefore, substituting these values in the formula and simplifying, we have:

$$Mode = 15.5 + \left(\frac{1}{1+5}\right) X\ 5$$
$$= 15.5 + \left(\frac{1}{6}\right) X\ 5$$
$$= 15.5 + 0.1667 X\ 5$$
$$= 15.5 + 0.8335$$
$$= 16.3335$$
$$= 16.3$$

Hence the mode is 16.3, rounded to 1 decimal place.

Measures of Central Tendency

I. Mean and Mode

The symbol for a population mean is μ.

The symbol for a sample mean is $\bar{x}$ (read "x bar").

Rounding Rule for the Mean: The mean should be rounded to one more decimal place than occurs in the raw data.

Example 1: Find the mean of 24, 28, 36

Mode

The mode is the value that occurs most often in a data set. A data set can have more
than one mode or no mode at all.

Example 2: Find the mode of 2.3 2.4 2.8 2.3 4.5 3.1.

Example 3: Find the mode of 3, 4, 7, 8, 11, 13.

The mode is the only measure of central tendency that can be used in finding the
most typical case when the data are categorical.

The procedure for finding the mean for grouped data uses the midpoints of the classes. The formula for finding the mean of grouped data is $\bar{x} = \frac{\sum f \cdot x_m}{n}$.

The modal class is the class with the largest frequency.

II. **Median and Midrange**

The median is the midpoint in a data set.

The symbol for a sample median is MD

1. Re-order the data from small to large
2. Find the data that represents the middle position

Example 1: Find the median

(a) 35, 48, 62, 32, 47

Solutions: re-arrange; 32,35,47,48,62 hence the media is 47, since it is the mid value after values were arranged.

(b) 25.4, 26.8, 27.3, 27.5, 28.1, 26.4

Solutions: re-arrange the values; 25.4, 26.4, {26.8, 27.3}, 27.5, 28.1. since there are six numbers (i.e even number of items) find the average of the 2 mid numbers i.e 26.8 and 27.3

Median = 26.8+27.3/2= 54.1/2 = 27.05 Approximately 2.7

Midrange is the sum of the lowest and highest values in a set of data, divided by 2.

Example 3: Find the midrange of 17, 16, 15, 13, 17, 12, 10. = 10+17= 27/2 = 13.5

Measure of Dispersion

Consist of Range, variance, and standard deviation

Range

Range is the highest value minus the lowest value.

R = highest value – lowest value

Example 1: Find the range of 32, 78, 54, 65, 89.

Variance and Standard Deviation

The measures of variance and standard deviation are used to determine the consistency

of a variable.

Variance is the square root of standard deviation, the distance that each value is from the mean. Hence it measures the dispersion away from the mean

Formulas for calculating variance and standard deviation

Definition Formulas

Variance of a sample $$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Standard Deviation of a sample $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$

$s = \sqrt{\textbf{variance}}$

Example 2: Use the formula to in table below find definition formula the variance and standard deviation of 5, 8, 11 $\bar{x} = 6$

x	$x - \bar{x}$	$(x - \bar{x})^2$
5	-1	1
8	2	4
11	5	25
		$\sum (x - \bar{x})^2 = 30$

$\bar{x}$= 5+8+11/3=24/3=6, Hence $\bar{x}$=6

Sample variance:

Sample standard deviation:

Example 3: Use the formula in table below to find the standard deviation of

5.8, 4.6, 5.3, 3.8, 6.0

$\bar{x}$ = 5.8+4.6+5.3+3.8+6.0/5 = 25.5/5 = 5.1

x	$x - \bar{x}$	$(x - \bar{x})^2$
5.8	0.7	**0.49**
4.6	-0.5	**0.25**
5.3	0.2	**0.4**
3.8	-1.3	**1.69**
6.0	0.9	**0.81**
		$\sum (x - \bar{x})^2$ **= 3.39**

Sample variance:

Sample standard deviation:

Example 4: Use the computational formula to find the standard deviation of

5.8, 4.6, 5.3, 3.8, 6.0

x	x^2
5.8	**33.64**

4.6	**21.16**
8.3	**28.09**
3-8	**14.44**
6.0	**36.0**

Note: Both the mean and standard deviation are sensitive to extreme observations

called the outliers. The standard deviation is used to describe variability when the mean is used as a measure of central tendency.

II. Variance and standard deviation for grouped data

The formula is similar to the computational formula of s^2 for a set of data

$$s^2 = \frac{\sum f \cdot x_m^2 - \left[\left(\sum f \cdot x_m\right)^2 / n\right]}{n-1}$$

Example 1:

The data below represent the net worth (in millions of dollars) of 50 businesses in a large city.

Compute the variance and standard deviation.

Class Limits	**10-19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**
Frequency	5	10	3	7	18	7

Solution

Class Limits	**10-19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**Sum**
Mid – Value (x)	15	25	35	45	55	65	
Frequency (f)	5	10	3	7	18	7	50
x^2	225	625	1225	2025	3025	4225	
Fx	75	250	105	315	990	455	2190
fx^2	5,602525	62,500	11035	99225	980100	207,025	6962410

$$s^2 = \frac{\sum f \cdot x_m^2 - \left[\left(\sum f \cdot x_m\right)^2 / n\right]}{n-1} = \frac{6{,}962{,}410 - ((2190)^2 \div 50)}{50-1}$$

$$= \frac{6{,}962{,}410 - (4{,}796{,}100 \div 50)}{49} = \frac{6{,}962{,}410 - 95{,}922}{49}$$

$$=\frac{6{,}866{,}488}{49} = 140{,}132.41$$

Standard Deviation = $\sqrt{s^2} = \sqrt{140{,}132.41}$ = 374.34

Mean = $\frac{\sum fx}{n} = \frac{2190}{50} = 4.38$

III. Coefficient of variation

The coefficient of variation is a measure of relative variability that expresses standard deviation as a percentage of the mean.

$$CVar = \frac{S}{x} \cdot 100\% = \frac{374.34}{4.38} \times 100\% = 8{,}546.58$$

When comparing the standard deviations of two different variables, the coefficient of variations is used.

Conclusion

The paper carefully examined methods of data collection and type of Statistics. Examples of Inferential and descriptive Statistics given, process involved on how inferential is made about population based on analysis of sample selected from the population described with illustrations. Descriptive and inferential statistics are interrelated, so it is almost always necessary to use methods of descriptive statistics to organize and summarize the information obtained from a sample. While inferential can be applied in order to obtained an accurate vital analysis and meaningful conclusion of the subject under investigation without Statistical bias and acceptable conclusion.

References

Adeyemi, T.O. (2002). *Introductory statistics for educational research*. Ado-Ekiti: Green Line Publishers, Pp 28-47.

Adeyemi, T.O. (2009). Inferential statistics for social and behavioral research. *Research Journal of Mathematics and Statistics Vol. 1(2), Pp. 47-54.*

Ali, G. (2011). Social studies research methods and statistics. In Y. Kadiri, C.O. Ololobou, T.S.

Ahmed, A.G.A. Zuru (eds). *Fundamentals of social studies education*. Pp. 309-348. Kano: Jaleyemi Graphics & General Enterprises.

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative Research, Vol. 1, no. 3: pp. 385–405*. Retrieved from http://goo.gl/vpqeqj 02/10/18.

Baddie, E. & Halley, F. (1995). Advantages in *social research: Data analysis using SPSS for windows.* California: Pine Forge Press.

Basu, R. (2004). *Implementing quality- a practical guide to tools and techniques*. UK: Thomas Learning. PP 165-173.

Gill, P., Stewart, K., Treasure, E. and Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. *British Dental Journal Vol. 204, pp 291–295.*

Kothari, C.R. (2004). *Research methodology and techniques*. New Delhi: New Age International Limited Publishers. PP 238-272.

Kolawole, E.B. (2001). *Tests and Measurement*. Ado-Ekiti: Yemi Printing Services.

Ololobou, C.O. (2007). Social studies research methodology; In Y. Kadiri, C.O. Ololobou, N. Celestin (eds). *Social studies for tertiary institutions.* Pp. 178-196. Kano: Tahir Advertising Agency.

O'Leary, Z. (2014). *The essential guide to doing your research project*. Thousand Oaks, California: SAGE Publications Inc. PP 87-110.

Queiroz, T., Monteiro, C., Carvalho, L., & François, K. (2017). Interpretation of Statistical Data: The Importance of Affective Expressions. *Statistics Education Research Journal*, *16*(1).

Salant, P. and Dillman, D.A. (1994). *How to conduct your own survey*. New York: Wiley. PP 77-95.

Silverman, D. (2000). *Doing qualitative research*. London: SAGE Publication. PP 27-35

Songsore, E., & White, B. J. (2018). Students' perceptions of the future relevance of statistics after completing an online introductory statistics course. *Statistics Education Research Journal*, *17*(2), 120-140.

Verma, J. P. (2019). Importance of Statistics in Psychology. In *Statistics and Research Methods in Psychology with Excel* (pp. 1-21). Springer, Singapore.

Weiss, C.H. (1999). The Interface between evaluation and public policy. *Evaluation Journal Vol.5(4) Pp. 468-486*