Kelechi Joseph Dilaku
UM82041HE91258

COURSE NAME:
(Health Science)

Assignment Title:
(Statistics in Health Science)

ATLANTIC INTERNATIONAL UNIVERSITY
October/year

Table of Contents

1. Introduction

The application of statistics in health science is not just theoretical. Statistics in health science is concern with the application of statistical techniques and methods, in collecting, analyzing, interpreting, and guide in making proper decision in presenting data related to health science. Also “Health statistics are numbers that summarize information related to health” (MedlinePlus, 2020). Statistics in health science is an arm of statistics that applies computational method, and mathematical tools to help in solving health science related issues.

As a good instrument to evaluate the effectiveness and safety of health issues. Statistics as a discipline that helps health researchers to understand the causative factors, patterns, trends, distribution of diseases in a population, and safety of interventions is to be considered. The processes of collecting data, visualizing, summarizing, relationships, patterns, and trends in data analysis is not excluded. Not forgetting the important phase of data collection from various sources to respond to paramount research questions, test and evaluate result. The inferential statistics such as ANOVA and regression. The important role of statistical methods in analyzing data and decision making. Statistics in health science being very important in enhancing scientific knowledge, and an effective health outcome in health science fields such as biostatics, public health, health informatics and epidemiology is not forgotten.

Statistic in health science is the implementation of statistical methods to collect, analyze and interpret health related data. Statistical methods and techniques are good tools for health research, data analysis and decision making. This claim is supported by the following researched work on the various topics, such as the concept and principle of statistics, concept of data collection and description, and concept of data analysis and inference.

1.1 Section A Concept and principle of statistics

To fulfil the purpose of data collection and analysis, statistics has to be considered. The concept and principle of statistics can be summarized as a science that pertains to a collection of methods used in collecting, analyzing, explaining, and presenting outcome data. It has two main branches, which are descriptive statistics and inferential

statistics. It has some basic concepts as population, parameter, bias, variability, sample, measurement and others. The study looks into the concept and principle of statistics.

The concept and principle of statistics is a discipline that can make use of mathematical tools in data analysis. It is “the science of collecting, analyzing, presenting and interpreting data” (Williams, Sweeney, Dennis, Anderson, & David, 2023). The numerical values that describe a sample or collections of samples from which conclusions may be drawn is data. It is recognized in statistics that there are variabilities in the data, meaning that all things are not exactly the same. Statistics is an art and also a science of making decisions which depends on qualitative evidence.

Some basic concepts of statistics are as follows: Sample: This is any subset of the population, which are randomly selected. Population: This is any specific collection of objects of interest, such as chalk, plants, animal, and people. Measurement is a numerical value computed for each member of a population, such as weight, height, color and others. Statistic: This is a numerical value computed from the sample data which can be used to test hypothesis. Random variable: This is a variable which its value depends on chance, such as a dice roll, survey response and outcome of a coin toss.

Other basic concepts of statistics are as follows: Parameter: This is a numerical value that summarizes the expression of the population as a whole, such as the standard deviation, mean, median mode and others. Variability: This is the degree by which measurement varies from one sample to another. Bias: This is the systematic error from the true parameter or inaccurate representation of the population. Validity: This is the degree to which a measurement is accurate to the real world. Reliability: This is the degree to which a measurement is consistent and was able to reproduce the same result under similar conditions.

Statistics is a useful tool that can help in making a reasonable and sensible decision based on real evidence. Application of mathematical concept, and methods to statistics helps one to develop problem-solving skills to real world situations. It helps to understand and interpret data in various discipline such as health, business, education, social science and science. Some common statistical tests used in health science are as follows: One-way ANOVA, Two-way ANOVA, Student’s t-test, Chi-square test, Mann-Whitney U test, Regression analysis and Correlation coefficient test.

To properly make use of statistics involves carefulness, critical thinking, and ethical considerations. If care is not taking statistics can be misused, by producing a manipulated result, which will be misleading. Therefore, one ought to be conscious of the limitations and ascertain the validity of the data used.

Statistics has two main categories, which are descriptive and inferential statistics. It is simply put as “while descriptive statistics summarize the characteristics of a data set, inferential statistics help you come to conclusions and make predictions based on your data” (Bhandari, 2023). Descriptive statistics as one of the categories of statistics

involves describing, organizing and displaying of data. It summarizes data in a clear and understandable way using numerical values, tables, graphs, charts and other methods.

There are measurement of position and three main types of descriptive statistics, such as the central tendency which has to do with the average of the values. Examples are mean, media and mode. The dispersion or variability which has to do with the spread out of the values. Examples are interquartile range, range, standard deviation and variance. The frequency distribution which has to do with the distribution of values. Examples are count, percentage, frequency, tables, and graphs. Measurement of position which has to do with location. Examples are quartile and interquartile.

Another main category of statistics is inferential which involves arriving to conclusions with a rational inference about populations based on samples. It can use methods such as confidence intervals, regression analysis, and hypothesis testing to make data-based predictions. The descriptive statistics try to summarize the characteristics of a data set while inferential statistics gives opportunity to test a hypothesis. This is to test whether a data is generalizable to the broader population.

My opinion and analysis of the concept and principle of statistics, is that statistics is a discipline used for collection, and analyzation of data, to identify common patterns, and trends useful for decision making, in health institute and other organizations. It does that, by converting the pattern and trend into meaningful information, which can help organization optimize processes, increase profits and reduce costs. For statistics result not to be bias, it requires appropriate data quality, reliability and validity
.
Another opinion and analysis of the concept and principle of statistics is that, it helps researchers design experiments, summarize results from collected data and draw conclusions. It is a useful tool for scientific inquiry, which helps researchers to test casual claims and measure relationships. In other to avoid misleading result, statistics analysis requires good planning, interpretation and effective communication.

I would apply the knowledge of the concept and principle of statistics in my life, work and community as follows:
In my life, it would help me to measure my personal health, emotion and general well-being. I would do that by using statistics to track my sleep quality, mood, and weight. I would use it to monitor the progress of my set goals. Statistics can also help me to be organized by producing a timetable for my undertakings such as, how to manage stress, how to exercise, what to eat, when to eat and so on.

In my work, the concept and principle of statistics would help me as an epidemiologist to improve efficiency, performance, quality and productivity. I would achieve that with the concept of statistic, by using it in my data collection, analysis and result presentation. It would also help me to present my results and findings to my coworkers in a simple, convincing, and clearer manner.

In my community, I would use statistics in attending to social services such as health care, education, and environment. I would do that by using data collected with the help of statistics from various sources, such as media, census and survey. I would use statistics to summarize and display my findings in various ways, such as graph, table and chart. I would also use statistics to monitor patterns, trends, and correlation between any outbreak events and its causatives in my community.

My personal experience in the study, the concept and principle of statistics is in the following: I design an experiment to ascertain how people in my community use insecticides-treated bed net. My intention was to know the average number of nursing mother that uses insecticides-treated bed net, and how it varies when compare to those that are not nursing. It became imperative that I have to use statistics to help me answer these questions.

To start with, I define population of interest as all the nursing mother that are living not too far from my house. To define population “in statistics, population is the pool of individuals from which a statistical sample is drawn for a study” (Momoh, Murry, & Rathburn, 2023). Having define population of interest, I designed a survey that focuses on knowing gender and age bracket of babies, which their mothers like to use insecticides-treated bed net. I randomly selected 50 nursing mothers, from the population, and focused my survey on them.

I judiciously, collected and organized my data from the survey. I made use of descriptive statistics in determining the standard deviation, range, mean, media, and mode from my sample data. I didn’t stop there, but used the sample data to create frequency tables, and histograms based on the distribution by age of the babies.

Finally, based on the sample data, I made used of inferential statistics to make generalizations about the population. With the use of confidence intervals and hypothesis test, I was able to estimate, and test the population mean of the number of nursing mothers that uses insecticides-treated bed net, based on their babies age. I also used ANOVA, and chi-square tests to compare the means, and proportions of the number of nursing mothers that uses insecticides-treated bed net, based on their baby’s age. For more understanding, “The Chi-Square test is a statistical procedure for determining the difference between observed and expected data” (Biswa, 2023).

With the above experiment using statistics, I was able to understand the importance, nursing mothers attached to the use of insecticides-treated bed net for babies in my community. I understood the sources of variability in my data, and the way to present them in my analysis. To interpret data in meaningful way is one of the use of statistics.

The case example to demonstrate the concept and principle of statistics, is how I would use statistics in the test of the effectiveness of new drug for the treatment of malaria, as a pharmacoepidemiologist. I would neither give the new drug to every person who has malaria, nor would I withhold it from those who need it. In facts, I would use statistics to

design an experiment, where I would randomly assign some patients to receive the new drug as group (A), and some to receive a placebo as group (B), which serves as the control. I would compare the outcomes of the two groups, and use statistics to determine if there is meaningful difference or result in the effectiveness of the drug versus the placebo. The difference in this experiment is a statistic that measures the effect size of the use of the new drug in treatment. I can also use statistics to quantify how likely it is that my result is due to chance, and how I can generalize it to the population of all malaria patients.

.

The picture, table and graph originally produced by me to illustrate the concept and principle of statistics using 150 dpi resolution is shown below:

Figure: 1. The picture below is used to demonstrate the use of data set in summarizing a set of observations of the use, and effect of drugs in a population.

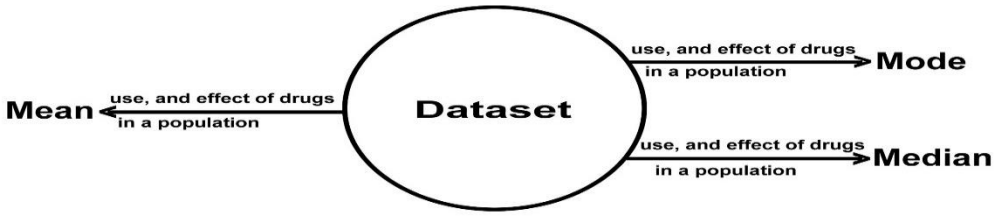
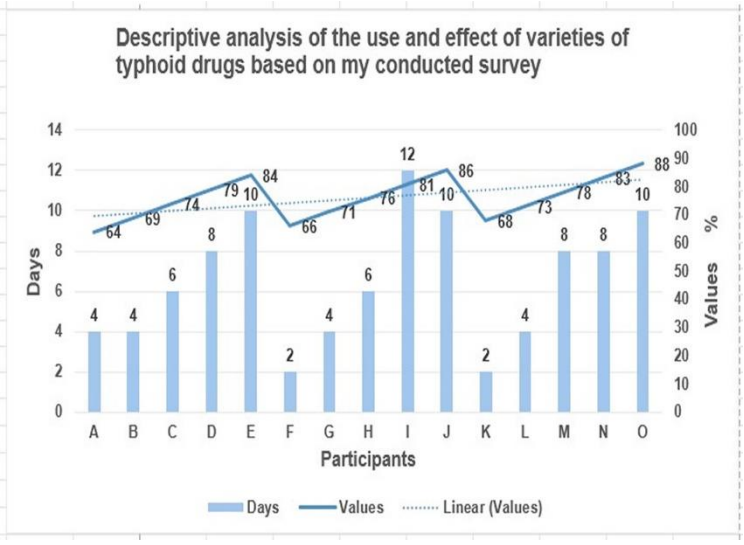


Fig 1.1 is a table for the graph of fig 1.2 below.
Fig 1.2 is a descriptive analysis graph that display the distribution of the number of days of using varieties brands of typhoid fever drug, in Otoro community, and the estimated values of the use, and effect of drugs on each of the 15 participants in percentage.

Participants	Days	Values
A	4	64
B	4	69
C	6	74
D	8	79
E	10	84
F	2	66
G	4	71
H	6	76
I	12	81
J	10	86
K	2	68
L	4	73
M	8	78
N	8	83
O	10	88

Mean	6.53333
Median	6
Mode	4
Range	10
SD	3.15926



The following is the statistic of my dataset:
The mean number of the number of days of using a brand of typhoid fever drug is 6.5.
The median number of days of using a brand of typhoid fever drug is 6.
The mode number of days of using a brand of typhoid fever drug is 4.
The range number of days of using a brand of typhoid fever drug is 10.
The standard deviation of number of days of using a brand of typhoid fever drug is 3.2.

The mean percentage is 45%.
The median percentage is 40%.
The mode percentage is 30%.
The range percentage is 70%.
The standard deviation percentage is 20.2%.
You can see that each point on the graph correspond to a row in the table. For instance, the point (10,88) on the graph corresponds participant O, who use a brands of typhoid fever drug for 10 days and has 86 values in percentage.
The dotted line on my graph shows the general direction of my set of data points on the graph. This trend line helps to understand how data changes over time.

1.2 Section B Concept of data collection and description

The gathering and evaluating information before use is paramount. Data collection is an organized procedure of collecting and analyzing information from various sources in other to respond to research questions, test hypotheses, and evaluate the results. Different methods of data collection are surveys, experiments, observation, interviews, and secondary data analysis. There are types of data that can be collected which are qualitative and quantitative data. The study looks into the concept of data collection and description.

Data collection is an organized procedure of gathering and analyzing information from various or multiple sources in other to respond to research questions, test hypotheses, and evaluate the results, while data description is the representation of features, and patterns of collected data. Also “data collection or data gathering is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes” (WIKIPEDIA, 2021). Data collection is an important phase of all forms of research, decision making and analysis. Accurate data collection is very important to make informed healthcare decisions, keep research integrity, and ensure quality assurance.

Data collection concepts are as follows: data, data types, data sources, data collection methods, and data collection tools. Data collection tools need to be chosen in consideration to the research objectives, quality, and accessibility of data source, the time frame, and budgets. Data collection tools: These are instruments that are used to store, process, and collect data. Tests, checklist, audio recorder, video camera,

sensors, software application, and questionnaires are some of the examples of data collection tools.

The two types of data collection methods are qualitative and quantitative. Qualitative data is analyzed through categorizations and interpretations. It is expressed in words, themes, and categories to make a meaningful interpretation of data. Quantitative data is analyzed through statistical method to show the distribution, frequency, and is expressed in numbers and graphs. For more understanding, “qualitative data collection refers to non-numerical research that gathers information on concepts, thoughts or experiences” (Indeed Editorial Team, 2023). Quantitative data is the reverse or opposite of qualitative, it is numerical or statistical information incline.

Data collection can be done using different methods, such as survey, focus group, experiment, observation, ethnography, interview, and secondary data analysis.
Survey: This is the use of set of questionnaires to extract information from large number of participants. It is a useful method for collecting quantitative or qualitative data like behaviors, attitudes, characteristics, and opinions of a population

Focus group: This is a small group of number of participants that are facilitated to share their views and opinions on a topic. It is a type of qualitative research. It can provide comprehensive insight into the feelings, thoughts and views of the participants. Focus group is useful in exploring preference, opinions, attitude, and behaviors.

Experiment: This is a type of data collection that is used to test variables relationship and measurement of their effects on other variables. It is often used in qualitative research to test hypotheses. In this data collection method, the researcher tries to affect the independent variables and measures their effects on the dependent variable. The following factors that may influence their result are to be kept in check, such as blocking, randomization, and binding. It is used in a clinical trial, to understand the effectiveness of a new drug in a population.

Observation: This involved systematic recording of interest, either the character of living beings, objects, phenomena, or control setting. Depending whether the data collected are numerical or non-numerical this method can be either quantitative or qualitative. Depending on how data are collected observation can be primary if the data was collected directly by the researcher, and secondary if is from previous research.

Ethnography: This is a type of qualitative research that focuses on studying the way of life of a group of people, in their natural setting, in the area of their culture and behavior. “Ethnography is a qualitative research method in which a researcher- an ethnographer- studies a particular social/cultural group with the aim to better understand it” (Kramer & Adams, 2017). Ethnographers uses the following method of data collection to gain more understanding of the group of interest such as interview, participant observation, surveys, archival research and naturalism. Ethnographers can adopt the use of these

aforementioned methods in their research questions, objectives, and resources gathering.

Interview: This method involves gathering the opinion, feelings, belief, stories, and experiences of people conducting conversations with either an individual or the members of a group. Depending on the direction by the interviewer or researcher, the interview can be unstructured, semi-structured or structured.

Secondary data analysis: This is a method of using the data that was collected by another person for a different reason. Sources of secondary data are as follows: reports, newspaper, online databases, historical archives, commercial organizations, published books, journals, government agencies, and academic institutions.

Having known that the data collection is the basis of any research work, the accuracy of the data determines the reliability and validity of the findings. For accurate and appropriate data collection, researchers should ensure that they do the following: Define and specify their objectives of the research. Choose appropriate data sources that is suitable for that type of research. Test and evaluate the data collection instrument for reliability. Analyze the collected data using the appropriate statistical techniques, communicate and report the findings.

It is good at this juncture, to throw some light into data description instead of dwelling on data collection. Data descriptions: These are the representations used to understand the main features and patterns of the collected data. These representations can be done using the type, name, size, and meaning of each variable in a collection of data or dataset. These representations can be well expressed, using the various techniques, such as, measurement of central tendency, measurement of variability, chart, histogram, frequency distribution, table, graph, and others. Different types of data descriptions are as follows: data dictionary, descriptive statistic and data type.

My opinion and analysis of concept of data collection and description, is that it is a complex topic and encompasses different aspect of research methodology, data presentation, and data analysis. It helps one to make a choice of data collection methods and appropriate tool, depending on the type of research, the aim of the study and the view of researcher. It is an essential stage of research, decision-making and analysis, which can be applied in health research.

I would apply the knowledge of the concept of data collection and description in my life, work and community as follows:
In my life, I would use the concept of data collection and descriptions in making good decisions, such as selecting a health insurance plan, and choosing a career.

In my work, the concept of data collection and description would help me to improve my productivity, performance and quality. As an epidemiologist, I would use data analytic tools, such as R software or Excel software to organize, transform, and interpret the

data I collected from a population. I would use data presentation tools, such as Google slides, Prezi and power Point to create reports and recommendations.
I would also use data mining techniques, such as association, clustering, and classification, to detect trends, insights and patterns from the data.

In my community, I would use the concept of data collection and description in attending to social issues, such as health, and education. I would use data tools such as visuals and narratives to pass message across to my audience to take action about the benefit, and risk of drug use.

My personal experience in the study, concept of data collection and descriptions, is a survey conducted by a pharmacoepidemiologist, to understand the risk, and benefits of a new drug in a vast population. The pharmacoepidemiologist, design a questionnaire that ask the new drug users to rate various aspects of their experiences, such as it effectiveness to their reason of taking the new drug. He also open-ended questions that allow the drug users to express their opinions, and suggestions in their own words. The data collected from the survey is quantitative, if is considered as ratings, and qualitative, if is considered as comments. The data description included information such as the sampling method, the margin error, the number of respondents, the main findings and the consensus result from the data analysis. The pharmacoepidemiologist used the concept of data collection, and description to identify the risk, and benefit of the new drug in a vast population.

The case example to demonstrate, the concept of data collection and descriptions is as follows: A master student of epidemiology want to understand the relationship between distribution, and determinants of health-related states in a specified population. He decided to collect quantitative data, by using a survey method. He designs a questionnaire that ask the participants, to rate their emotion levels concerning mosquito bite, and the result of those that are exposed to it, in a scale of 1 to 10. He distributed the questionnaire in a community to a random sample of 100 people and their responses were collected. In other to describe his data, he used descriptive statistics such as standard deviation, mean, median, mode, and correlation coefficient. He also used graph such as scatter plots, and histograms. He further makes use of box plots to visualize the data and identify any outliers and patterns. “Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand” (Brush & Burns, 2022).

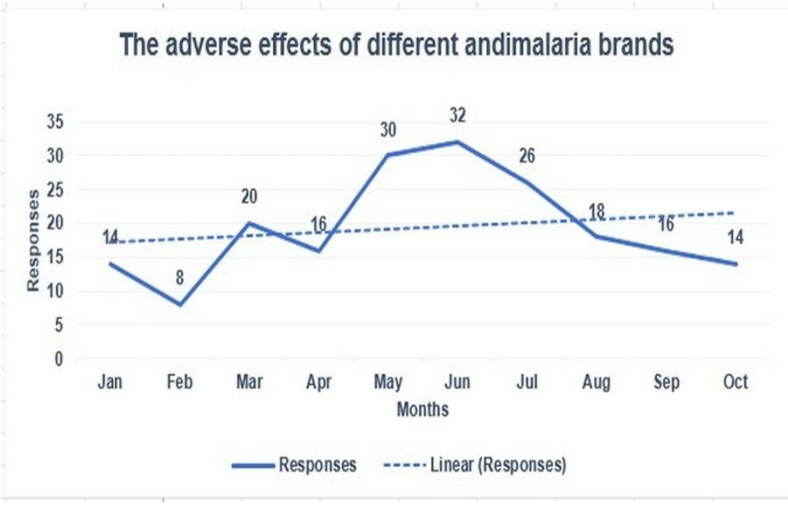
The picture, table and graph originally produced by me to illustrate the, concept of data collection and description using 150 dpi resolution is shown below:

Figure: 2. The picture below is used to demonstrate the collection of responses to a survey, which forms my dataset.

Figure: 2.1 is the table of the graph of fig 2.2. Fig 2.2 is a graph that shows the outcome of the distributed survey in Otoro community, which asked 100 participants, the brand of antimalarial drug they took, their adverse effect, and the month they took they drug based on their responses in percentage.

Months	Responses
Jan	14
Feb	8
Mar	20
Apr	16
May	30
Jun	32
Jul	26
Aug	18
Sep	16
Oct	14

Mean	19.4
Median	17
Mode	24



The following is the statistic of my dataset:
The mean response is 19.4 %.
The median response is 17 %.
The mode response is 24 %.
The dotted line on my graph shows the general direction of my set of data points on the graph. This trend line helps to understand how two variables are related to each other.

1.3 Section C Concept of data analysis and inference

In other to be able to make prediction from data, understanding should be the first approach. The concept of data analysis and inference can be seen as the two important aspects of statistics that aid one to understand, and be able to make predictions from data. Some of the concepts and methods involved in data analysis and inference are: Descriptive statistics, predictive analysis, and others. The study looks into the concept of data analysis and inference.

The concept of data analysis and inference can be described as ways to understand and make a comprehensive prediction from data. Data analysis is the steps or processes of visualizing, summarizing, and a way of exploring data, to find patterns, relationships, and trends, while data inference remains the process of using data analysis to make decision, and come to conclusion about a vast population based on a sample of data. It is “the process of collecting, modeling, and analyzing data using various statistical and logical methods and techniques” (Calzon, 2023). Data analysis is descriptive, while data inference is inferential. A descriptive statistic is used to describe and organize data. Inferential statistics on the other hand try to use analytical tools to draw conclusions concerning population from samples. Inferential statistics “is used to analyze the probabilities, assumptions, and outcomes of a hypothesis” (Hassan, 2023).

Some of the method of data analysis, depending on the type, purpose, and nature of data are as follows: Text analysis: This method of data analysis involves extracting information from unstructured text, such as email, review, documents, social media posts and others. It is a type of qualitative data analysis. It can be used for the purpose, such as summarization, topic modeling, sentiment analysis, keyword extraction and others. Descriptive statistics: These are numerical summaries of data that helps to describe central tendency, population and variability of data. Examples are: Graphical representations such as boxplot, histogram, and scatterplots; Measure of dispersions such as standard deviation, range, variance, and interquartile range; measure of central tendency such as mean median and mode.

Data mining: This is the method of data analysis involves trends, patterns, and relationships in vast and very complex data set. Data mining is also present as “the process that turns source data into something useful” (Whitmore, 2023). It is a type of quantitative data analysis and can be used for the purpose such as anomaly detection, classification, association rule mining, and clustering. Cluster analysis: This is the method of data analysis that involves arranging similar data points into clusters depending on their attributes. It is a type of mathematical data analysis. It can be used in different purposes such as outlier detection, segmentation, and others.

Predictive analysis: This is the method of data analysis that involves the use of current and historical data to forecast future outcomes. It is obvious that, descriptive analytics or analysis “looks to what has happened in the past “ (Stevens, 2023). It is a type of mathematical data analysis. This method of data analysis can be use in various purposes, such as prediction of disease or event outcome, risk, and others. Diagnostic

analysis: This is the method of data analysis that involves detecting the factors that influenced an outcome. It can be used for problem solving, casual inference and others.

Machine learning: This is seen as a branch of artificial intelligence which uses data and algorithms to learn and perform task that only human supposed to perform such as clustering, classification, and regression. It can be used to unveil complex relationship and patterns in a data that are not feasible using traditional statistical methods. Inferential statistics: These are methods such as hypothesis testing, estimation, p-values to quantify the uncertainty and variability in the data, and the conclusions that make use of data from a sample, to make prediction about a population.

Data analysis is to data inference what foundation is to a building. Data analysis provides the basis for data inference, this is because it gives clear understanding of data in other for one to be able to make claims and decisions about it. It is “an aspect of data science and data analytics that is all about analyzing data for different kinds of purposes” (Urwin, 2023). The use of the analogy of detective can be implemented to understand the relationship between data analysis and data inference. While data analysis is collecting useful data information and finding patterns in the evidence, the data inference is arriving into conclusions and making logical reasoning based on the useful data information.

Data inference being a process of making predictions about a population which is all based on a sample data taken from that population. Different types of data inferences are used for different data types and purposes. Examples are: Bivariate regression, Chi-square test, confidence interval, contingency table, hypothesis testing, multivariate regression, and personal correlation.

There are several tools that can be used in data inference depending on the purpose or type of the analysis. Some common tools used in data inference are: Statistical methodology: This is set of procedures and techniques that are used to make references about the population and analyze the sample data. It includes choosing the appropriate models, statistical tests, and assumption for the data.

Inference servers: These servers work as a help to deploy and manage task such as scaling, load balancing, monitoring, logging and machine learning models in production environments. Machine learning model: These model uses algorithms that learn from data and improve with experience to make prediction based on input data. Examples are decision tree, support vector machines and neutral network.

Statistical inference languages: These are languages that helps to perform data analysis and inference with the use of various libraries and packages that support machine learning models and statistical method. Examples of these statistical software tools are Python, R, Excel, SPSS, MySQL, Stata, and SAS. Statistical software tools are those tools or applications that help to perform statistical analysis on data. Statistical

analysis, such as data collection, data cleaning, data visualization, data interpretation, and data modeling.

Visualization tools: These are tools that help to make a meaningful presentation and communicate the results of data inference in clearer ways. Examples are: graphs, charts, dashboards and maps. Data management tools: These are those tools that help to organize, store, and access the data destinations and sources for data inference. Examples are: Data warehouse, databases, and data lakes.

My opinion and analysis of the concept of data analysis and inference is how to avoid bias, in the interpretation of the results. This is because there are many ways one can be misled by the data, such as, selection bias, confirmation bias, confounding and overfitting. I therefore, infer that if such biases, and confounding must be rule over, the need to use standard, and logical evidence to evaluate the reliability, and validness of our predictions and conclusions has to be most considered.

I would apply the knowledge of the concept of data analysis and inference in my life, work, and community as follows:
In my life, I would use the concept of data analysis and inference to make a reasonable decision about my health, by using an authentic inferential statistic to compare different people opinions of certain medication.

In my work, the concept of data analysis and inference would help me to use hypothesis test, in evaluating the effectiveness of different health policies, and interventions. The concept would help me to improve my performance, and productivity in my work as an epidemiologist. This is by using it to compare the validity, and reliability of secondary data from different sources.

In my community I would use the concept of machine learning methods, to analyze vast health datasets, in other to discover patterns, trends that can help me to understand the occurrence, and distribution of Malaria, and Ebola in African countries. I would use such result to sensitize and promote health standard of my community.

My personal experience in the concept of data analysis and inference based on my interest and goal as a pharmacoepidemiologist, is my concern about the attitude of people towards varieties of drugs, I wish to use data to predict future health needs, to ascertain risk factors and evaluate interventions.

A case example to demonstrate the concept of data analysis and inference is as follows: An epidemiologist can use inferential statistics to make estimates, and test hypothesis about those factors that contribute to disease occurrence, transmission, frequency, and distribution of diseases in a population, based on sample data. He can make use of confidence intervals to estimate the proportion parameter, such as the percentage of events occurrence. He can also use hypothesis test, to draw conclusion

about population parameters, such as whether there is a noticeable correlation between the distribution of diseases, and potential factors of interest.

The picture, table and graph originally produced by me to illustrate the concept of data analysis and inference using 150 dpi resolution is shown below:

Figure: 3. The picture below is used to demonstrate a One-Way ANOVA test in Excel, to determine if there is a significant between the mean sales of antimalarial drug, by a pharmaceutical shop in three different local government areas, labeled Area 1, Area 2 and Area 3, which serves as the independent variable. The table for the test is shown in fig 3.2 below. The sales revenue in naira are the dependent variable. The three different areas are groups levels. In my analysis a One-Way ANOVA test was conducted to compare the means of the three groups. One-Way ANOVA "compares the means of three or more independent groups to determine if there is a statistically significant difference between the corresponding population means" (Zach, 2018). Having known that, the F-statistics and P-value was determined.

Anova: Single Factor

								k*(k-1)/2
SUMMARY								3
Groups	*Count*	*Sum*	*Average*	*Variance*				
Area 1	15	676	45.06666667	5.638095238				
Area 2	15	681	45.4	5.685714286				
Area 3	15	722	48.13333333	5.552380952				

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	84.9333	2	42.46666667	7.549097065	0.00158185	3.2199
Within Groups	236.267	42	5.625396825			
Total	321.2	44				

My p-value being 0.00158185 was very low, when compare with the alpha level, which is 0.05, hence, there is a significant difference in mean between my 3 groups. I then decided to perform a Post-hoc test as shown fig 3.1 below, to determine which group mean differ. To understand the exact, where these significant lies. Because the variances of my test, are approximately the same, I decided to use equal variance T-test. To know how many comparisons, I would have to do I use the equation below: n*(n-1)/2 where n is the number of groups 3*(3-1)/2 = 3*2/2 =6/2 =3. Therefore, I have three comparisons I have to make. This means, I need to compare Area1 vs Area 2, Area 1 vs Area 3, and Area 2 vs Area 3 respectively. I used separate T-tests and control the multiple comparisons with the Bonferroni method.

Because I do the same test repeatedly, I want to do a correction to account for the fact that I have increased my type 1 error rate by repeatedly doing the same test. This correction is Bonferroni correction. I apply this to my work, by dividing the original alpha

level 0.05, by the number of tests completed, which is 3. Mathematical expression: α/n=p. Where “α” is the original alpha level, “n” is the number of tests run, and “p” is the p-value for my comparison. Therefore, 0.5/3= 0.16666667.
This quotient I got from the division, is the p-value I used to compare each of my tests to determine whether any of the groups is significantly different from one another. From my result in fig 3.1, two groups comparison has a p-value for the two-tailed test that is less than 0.16666667. Therefore, I concluded that all my “Yes”, in fig 3.1 are the areas that have significant difference.

fig 3.1

POST-HOC TEST

AREAS	P-value (T test)	Significant?
Area 1 vs Area 2	0.704140245	No
Area 1 vs Area 3	0.001382148	Yes
Area 2 vs Area 3	0.003787483	Yes

ALPHA

Test	Alpha
ANOVA	0.05
Post-hoc test (Bonferroni correction	0.016666667

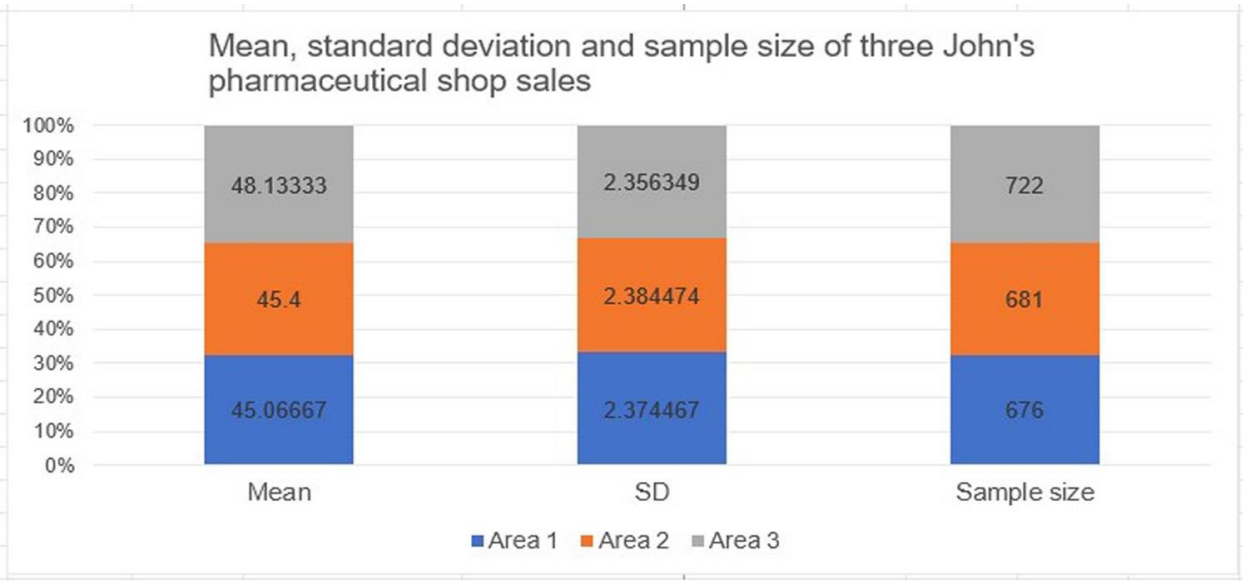
Fig 3.2 Below is the sales revenue data that was collected from John’s pharmaceutical shop, from three different local government areas over a specific time period in months. To ensure reliable analysis, I ensure that each pharmaceutical shop has sufficient sales record. It was used for the tests in fig 3 above.

	Area 1	Area 2	Area 3
	42	47	50
	50	42	43
	43	41	48
	42	46	49
	44	47	46
	43	44	47
	46	44	49
	48	47	50
	45	49	46
	45	47	50
	44	48	50
	45	47	45
	49	45	49
	45	42	52
	45	45	48
Mean	45.0667	45.4	48.133333
SD	2.37447	2.38447	2.3563491
Sample Size	676	681	722

Fig 3.3 below is the extrapolated table from fig 3.2, showing the average mean, standard deviation and sample sizes of the three groups.

Group	Mean	SD	Sample size
Area 1	45.0667	2.374467	676
Area 2	45.4	2.384474	681
Area 3	48.1333	2.356349	722

Fig 3.4 below is the visualization of fig 3.3 data and information above.



1.4 Conclusion

In sum, the application of statistics in health science is both theoretical and practical. This essay has looked into the title statistics in health science with the following topics: The concept and principle of statistics, concept of data collection and description, and concept of data analysis and inference, in statistics in health science, from different perspective and context. Such as the area of summary, description, analysis, application, personal experience and examples of each of the concepts in addition with pictures, tables and graphs.

The essay has shown that all this concept can be applied to day to day life. The essay has also shown that statistics in health science is a complex field that needs a lot of skill and understanding. It is a discipline that involves collecting data from its source, analyzing, interpreting and presenting data relating to health science. Statistics in health science can be of a help for an individual to understand how they can improve their health standard. The essay has highlighted some of the similarities and differences

among these perspectives and contexts, as well as some of the challenges and opportunities for collaboration and interpretation.

The research supports these meaningful good words: There is the need to be conscious of bias in data collection, and ascertain the validity of the data used, by always make use of the data sources that is accurate, well represented, complete, and generalized, to have an excellent, valid, and quality outcome. There is the need to have a strong team of data science which should include experts in machine learning, statistics, and visualization that will oversee and handle the vast and complexity of health science data sources. There is the need for a health statistics analyst to employ good planning, interpretation and effective communication, to avoid misleading result.

1.5 Bibliography

Bhandari, P. (2023, June 22). *Inferential Statistics | An Easy Introduction & Examples*. Retrieved from Scribbr: https://www.scribbr.com/statistics/inferential-statistics/

Biswa, A. (2023, October 11). *What is a Chi-Square Test? Formula, Examples & Application*. Retrieved from Simplilearn: https://www.simplilearn.com/tutorials/statistics-tutorial/chi-square-test

Brush, K., & Burns, E. (2022, December). *Data visualization*. Retrieved from TechTarget: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

Calzon, B. (2023, March 3). *Your Modern Business Guide To Data Analysis Methods And Techniques*. Retrieved from Datapine: https://www.datapine.com/blog/data-analysis-methods-and-techniques/

Hassan, M. (2023, October 1). *Inferential Statistics – Types, Methods and Examples*. Retrieved from ResearchMethod: https://researchmethod.net/inferential-statistics/

Indeed Editorial Team. (2023, October 13). *6 Methods of Data Collection (With Types and Examples)*. Retrieved from Indeed: https://www.indeed.com/career-advice/career-development/methods-of-data-collection

Kramer, M. W., & Adams, T. E. (2017). *Communication and Media Studies, Sociology*. Retrieved from SageResearchMethods: https://methods.sagepub.com/reference/the-sage-encyclopedia-of-communication-research-methods/i4910.xml

MedlinePlus. (2020, April 17). *Health Statistics*. Retrieved from medlineplus: https://medlineplus.gov/healthstatistics.html

Momoh, O., Murry, C., & Rathburn, P. (2023, April 29). *Population Definition in Statistics and How to Measure It*. Retrieved from Investopedia: https://www.investopedia.com/terms/p/population.asp#toc-what-is-population

Stevens, E. (2023, May 11). *The 4 Types of Data Analysis [Ultimate Guide]*. Retrieved from CAREERFOUNDRY: https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/

Urwin, M. (2023, June 15). *8 Types of Data Analysis*. Retrieved from Builtin: https://builtin.com/data-science/types-of-data-analysis

Whitmore, C. (2023, April 17). *Data mining: Definition, techniques, benefits, and risks in 2023*. Retrieved from NordVPN: https://nordvpn.com/blog/data-mining/

WIKIPEDIA. (2021, August). *Data collection*. Retrieved from wikipedia: https://en.wikipedia.org/wiki/Data_collection

Williams, T. A., Sweeney, Dennis, J., Anderson, & David, R. (2023, October 9). *statistics*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/science/statistics

Zach. (2018, December 27). *One-Way ANOVA: Definition, Formula, and Example*. Retrieved from Statology: https://www.statology.org/one-way-anova/