THE IMPACT OF STATISTICS

*ABDULMUHAMMID ABIOLA SANUSI, **ALIYUSUSMAN
*Federal University Kashere Gombe, Gombe state, Nigeria **Kaduna Polytechnic, Kaduna, Nigeria.

ABSTRACT
This research paper investigates the origin and development of statistics on science and technology, the role statistics have had on the representation of science over 20th century, the importance of statistics to every aspect of natural and social sciences. It also discusses the good statistics that had been collected to agreed good practices as a tool for the development of natural science. The contribution of some great men and women to the development of statistics, the steps involved in coalition of data, the statistical software available for data analysis and the use of statistics in analyzing data.

Keywords: Inception of Statistics, Science, Contribution of great men and women to Statistics, and Statistical Software.

Introduction
The statistical data collected and disseminated by the scientists and technologists about different findings are needed for a number of important purposes. They provided the information, the evidence, needed to draw conclusion on their findings both day to day research work and as well as making a reasonable decision on their findings. The statistical data gathered during the research work are also important for users outside the scientists and technologists. This information provided; gives the reality of their findings and also make individual to be informed about the ongoing findings and all its processes to achieve the reliable result, conclusion and decision.
Accurate statistics that has been collected according to agreed good practices; using the appropriate methods for data collection, processing, analyzing and draw conclusion are crucial as a tool for science and technology development in the world at all level of existence.
They provide an objective and replicable true picture of the condition and state of their findings; enable comparison, both over time and space, and set benchmarks for measuring progress in the future for further findings and research in the area of peculiarity and interest. Statistics provides the well-established data processed and analyzed using the statistics powerful tools for analysis, conclusion and decision making.

Statistics plays a vital role in every science and technology courses such as physics, chemistry, biology, agriculture, astronomy, geology, medicine and Information and computing technology (ICT) etc. therefore all science and technology courses have their specialized statistical package and software for accurate analysis of their data in order to disseminate accurate and reliable information about their related findings.

The value of statistics depends upon their quality. Since it is not easy to ascertain the quality of statistics directly, users must have confidence in the scientists, technologists and in the processes (i.e. the methods and standards) employed in the production of the statistical data. The knowledge of statistics guides the scientists and technologists to compile their data for future use and reference. It is the responsibility of the scientists and technologists to provide well established and comprehensive sets of data and also ensure that the data they provide are of good quality and that users are able to have confidence in the accuracy, reliability and integrity in the data; based on the knowledge they have acquired in statistics.

**Importance of Statistics**

In modern times, statistics is the back bone of decision making and its being introduced into all sectors of life; sciences, industrial, social, economic, political among others.

Statistics is no longer considered as a mere device for collecting numerical data but as a tool for developing and techniques for data handling, analysis and making viable inferences for them.

Statistical data is useful indispensable to planning. In order to ensure successful planning, correct analysis of accurate statistical data must be ensured.

Statistics has proved immensely important in solving varieties of prices, time-series analysis, index numbers and demand analysis.

Statistics is an indispensable tool in product marketing, inventory analysis and production control.

Statistics plays a vital role in the industrial sector, widely used in quality control, in production engineering as to know whether the product is conforming to specification or not.
Statistical data analysis has also proved useful in research and scientific like Biology, Astronomy, physics, Geography, Chemistry, and Medical Sciences.

THE ORIGINS OF STATISTICS

Historical Background
Statistics originated in the amalgamation of three currents of thought, each emerging in a different European country in the second half of the 1600’s: the German descriptive statistics, the British Political arithmetic and the French support for statistical surveys and thus their aim to determine their methodology. The word 'statistics' derived from the Latin 'status' or political state and the German 'statistik' – facts and figures for use of the state. Despite their earlier beginnings, statistics didn't become properly organized until the scientific revolution in Europe in the 17th century and the growth of European nationalism. Every nation then determined to gather as many facts as possible about its scientific achievement, technology development, population, trade, finances, taxes and armed forces. But it was not until the 20th century that official collection of statistics developed into the exhaustive process we see today, with data available on nearly every aspect of science, technology, and people's lives and on country's activities.

An indispensable tool for the new emerging science was provided by the probability theory, whose origins are undoubtedly French. It is worth taking a look at the individual role played by each branch in developing statistics as we know it today.

THE BIRTH AND DEVELOPMENT OF STATISTICS AND THE NEW NATURAL SCIENCE
As with all branches of science, statistics began as a practical activity, geared towards resolving practical everyday issues. Its exact origins have been lost in the mists of time, although it may be presumed that even the earliest societies needed to know the number of their individuals or families, how many could bear arms, how many heads of cattle, etc. Historical evidence points to this. It was only at a later stage that statistics was defined as a doctrine, in other words, was held to be a system with the aim of providing rules to perform practical tasks that is a method. Statistical techniques were first applied to matters of state, demographic questions and social phenomena and thus were largely responsible for shaping the theory and subsequent improvements. Thus, initially statistics was taken to mean both the technique as well as its application to demographic and social questions and State affairs.
Therefore, when tracing the birth and development of statistics its significance at the outset should be kept in mind, incorporating both a method and its application to the collective phenomena described, or in other words, the social sciences.

Statistics, in these terms, was born at a precise time: in the 17th century, at the very same time when, thanks to the efforts of Galileo Galilei (1564-1642) and Isaac Newton (1643-1727) and many other scientists, the new natural sciences and its main branches were also being forged. The same century also witnessed the invention of the 18th calculating machine and the bases were laid of modern mathematics, with the introduction of analytical geometry, mathematical analysis and probability theory.

This coincidence in timing, between the birth of statistics and the formation of the new natural sciences, was not merely accidental. In fact, the new natural science and statistics, as a method and a social science, while belonging to different worlds — natural phenomena, on the one hand, and matters of State and society, on the other, — shared a number of similar or common traits. The process leading to the birth and development of both sciences was based on the same factors and principles and triggered by similar needs. Each could emerge thanks to the 16th century appreciation of manmade tools created in the attempt to overcome man’s shortcomings. The new natural science was made possible by the invention and scientific use of instruments which went beyond man’s capabilities in their examination of nature.

Similarly, statistics as a method, by superseding human inability to quantify collective phenomena, permitted greater insight into these phenomena (originally those concerning the State and society). The new natural sciences and statistics followed the same approach, shared a mathematical basis, and pursued both scientific and practical aims.

FEATURES OF THE NEW NATURAL SCIENCES AND STATISTICS AND SIMILARITIES IN THEIR BIRTH AND DEVELOPMENT.

In tracing the birth and development of the new natural sciences and statistics we have come across a number of major scientists who were involved in both: Huygens, Descartes, Leibniz, etc. This is proof of the fact that both disciplines were children of the same intellectual environment. It was then common for observers who also dealt with different disciplines to transfer their methodological knowledge from one field to the other and, if possible, also any results matured.

The field of study of both the natural sciences and statistics was quite different: natural phenomena for the former and collective phenomena in
society and the State for the latter. Nonetheless, as outlined in paragraph 2, in their birth and development they shared a number of similar features.

1. They each sprang from the same intellectual background, which welcomed contributions from observers also operating in different fields and who were not necessarily professional scientists;

2. They were spurred by a new vision of the world (with regard to nature and society, respectively), which differed from that of the previous epoch, no longer viewed as a shadow of an ideal world but as a reality which had to be fully explored in total freedom and allow man to draw every possible advantage;

3. They were conceived of for practical purposes (the aim of natural science was to dominate nature for the benefit of mankind; that of statistics good government and the public interest) and thus adopted a functional approach to reality;

4. Their goal was to provide a quantitative description, offer an explanation for the phenomena being studied and pinpoint the underlying laws;

5. The inductive empirical method used was based on: (a) observing the facts, (b) the use of mathematics, on the basis of the facts observed, or rather experience, to elaborate the laws governing these phenomena (conceptual tools for the natural sciences were analytical geometry and infinitesimal analysis and, not until two centuries later, the probability theory, used from the very outset in statistics);

6. The collected quantitative data had to be exact to enable the application of mathematical procedures and thus the need for valid tools of observation;

7. They were made possible by the construction of new instruments that helped overcome man's shortcomings and thus enhanced his powers of observation. The birth of the natural sciences went hand in hand with the discovery of new tools compensating for mankind's limited senses (telescope, microscope, clocks, thermometers, hygrometers, etc.). Statistics originated with the devising of statistical technique which made up for the mind's inability to quantify collective phenomena (censuses, statistical surveys, etc. and the use of modules and questionnaires to collect data and tables and figures to represent these);

8. They gave rise to various disciplines. Offshoots of the new natural sciences included physics, in all its branches (mechanics, acoustics, optics, etc.), astronomy, chemistry, biology, etc. Statistics gave shape to statistics as we know it today, demography, social statistics, economic statistics, actuarial sciences, political economy, biometrics, anthropometrics, etc.
THE BIRTH AND DEVELOPMENT OF TECHNOLOGY USING STATISTICAL COMPUTING

The statistical computing began during the 1920s and 1930s as early IBM mechanical punched card tabulators are required by the universities and research labs. These machines are not only used for tabulating and computing summary of statistics but also for designing and solving complicated statistical models such as analysis of variance, linear regression and correlation. The most interesting is that, this knowledge of statistical computing has been extended to the invention of the large mainframe computer during the 1960s and the personal computer i.e. the desktop and laptop etc. in the 1980s.

In the early 1920s the labs created for the statistical computing really helped to advance statistical methodology. They helped to improve the idea of Galton’s and Pearson’s on correlation as a practical tool that could be used-for scientific research. This makes it necessary for researcher to think in terms of large problems with extensive datasets. With their ideology in statistical computing, modern statistical methodology could have remain interesting theory, useful for small problems but otherwise not in practice.

These labs also helped to advance the computation of science in general. Many of these labs offered their services to physicists and astronomers as well as to biologists and social scientists. Some created tables of higher mathematical functions. Other solved complicated differential equations. Most of these labs are located in Iowa State University and Columbia University, which became the test beds for early computer scientists, who experimented with new ideas for computing machines and for numerical algorithm.

The notable men of science with these early labs are familiar to very few — James Glover, H. T. Davis, A. E. Brandt, and Howard Tolley. They published little and made only marginal contributions to the theory of statistics or the development of computers. Yet these researchers held a deep faith that the combination of computing technology and mathematical statistics would radically change science. These tools remain linked today, but during the 1920s and 1930s the combination not only helped establish the field of statistics on the American continent, it also promoted computing as an important tool for scientific research.

The old statistical labs were probably more important to science and technology. In the 1920s, the computing labs helped establish statistics on the
American continent. Without them, even a modest study was beyond the ability of an individual statistician. At the same time, statistics labs often had the most powerful computing machines within their larger institution. They showed how organized computing could benefit science and provided a place for the earliest of computer scientists to test their ideas, though modern computers are more powerful than the antiquated technology of punch-card tabulators.

DEVELOPING AND MAINTAINING A GOOD STATISTICAL SYSTEM

Measuring has become an "Industry". Governments and their statistical offices have conducted regular surveys of resources devoted to research and development (R & D) since the 1960s. The methodology used is that suggested and conventionalized by the OECD Frascati manual, adopted by member countries in 1963 and now in its sixth edition, OECD(1962) since the 1990s, national government have also conducted regular surveys on innovation, again based on an OECD methodology known as the Oslo Manual. OECD/Eurostat (1997). More recently score boards of indicators on Science, Technology and Innovation, thereafter science. OECD(1995 & 2001)

The statistics collected by Official organization are regularly used by academics, among them economists who, over the last five decades, have produced a voluminous literature on measuring the contribution of science to economic growth and productivity (Z. Griliches 1998). Academics are also producers of their own statistics. Using scientific paper-counts as a tool, sociologists and others have studied the 'Productivity' of scientists since the early 1900s. Today, a whole community of researchers concerned with counting papers and citations call themselves bibliometrics. (B. Godin, 2006).

Before the 1920s, it was scientists themselves who conducted measurements on science. The statistics collected concerned men of science, or scientist: their demography and geography, their productivity and performance. This kind of statistics owes its development to the context of the time: measuring the contribution of great men, among them scientists, to civilization; then improving the social conditions of the scientists.

Starting in the 1940s, the kind of statistics collected changed completely, it was no longer scientists who collected them, but governments and their statistical bureaus. The most cherished indicator was thereafter money devoted to research and development (R&D). Again, this owes its development to the context of the time, namely science policy and efficiency. Science policy developed primarily due to concerns about using accounting as a way of controlling (government) expenses on R & D. But second, official
statistics also developed for a more positive aim: to determine target levels for the investment in scientific activities for public goods. OECD (1962).

CONTRIBUTION OF SOME GREAT MEN AND WOMEN TO STATISTICS
Thomas Bayes: (1701-1761): He was an English mathematician and probabilistic. During his life time, he published a few mathematical papers, of which the best – known is the theorem that nowadays, bear his name. The theorem is of basis important because it provides a solution to the general problem of inference and indication. The eponymous theorem has led to the development of an approach to statistics that run parallel to the method of hypothesis testing. This approach is referred to as Bayesian inference and its advocates are referred to as Bayesians.
Sir Francis Galton (1822-1911): An English doctor, explorer, meteorologist, biometrician and statistician. He used the correlation in its statistical sense. He made great use of the normal distribution and illustrated it in a lecture to the Royal Institution in 1874 using a quincunx. Galton’s idea on regression and correlation were promptly taken up and given a formal mathematical development by Karl Pearson and George Udny Yule.
Florence David Nightingale: (1820-1910): She was an English statistician pioneer of data analysis and graphical representations. She wrote that: Statistics is the most important science in the whole world.
Gregor Johann Mendel (1822-1884): He is an Austrian geneticist and statistician. He was definitely recognized as the originator of genetics only in 1930’s. He was the first to apply statistical methods in biology. Between 1856 and 1863, Mendel studied the hybridization of peas, first testing 34 of their varieties for the constancy of traits, and selecting 22 of them. He always examined a large number of plants to eliminate “chance of effects” and thoroughly planned his experiments.
Andrei Andreevich Markov (1856-1922): A Russian mathematician. His outstanding contribution was the idea underlying Markov processes. He is also remembered for his contribution to the mathematics of linear models through the Gauss Markov theorem. Markov, with Liapunov a discipline of Chebyshev, gave rigorous proofs of the central limit Theorem. Through his work on Markov chains, the concept of Markovian dependence pervades modern theory and application of random processes. His text book influenced the development of probability and statistics internationally.
Karl Pearson (1857-1936): An English mathematician, biometrician and statistician (1857-1956). He made prolific contributions to statistics and was one of the principal architects of the modern theory of mathematical statistics. In 1893, he started publishing articles on statistics. During the
period of 1895 to 1898 he presented a sequence of papers on correlation and in 1900 he proposed the chi-square test. Pearson's writing were prodigious; he published more than 650 papers in his life time of which 400 are statistical. Over a period of 28 years he founded and edited six journals.

Charles Edward Spearman (1863-1945): An English psychologist and self-taught statistician. His first two papers, introducing the Spearman's rank-order correlation and laying the foundation for factor Analysis, appeared in 1904. He is known for this two major contributions to statistics. His name is associated eponymously with spearman's rank order correlation coefficient.

George Udny Yule (1871-1951): He is a Scottish statistician. Yule's first paper on statistics appeared in 1895. In 1911 he published Introduction to the Theory of Statistics for the net forty years. During his time at Cambridge, he worked on the association in contingency tables, genetics and the theory of time series; introducing the terms correlogram and autoregressive models.

William Sealy Gosset (1876-1937): An English statistician. He was a pioneer in statistical theory associated with small sample sizes, and the student's t-distribution. These contributions placed Gosset among the men of the newly emerging field of statistical methodology. In fact, the t-test is perhaps the single most widely used statistical tool in applications.

George Waddel Snedecor (1881-1974): He is an American biometrician and Statistician. He is best remembered as a pioneer in making statistical tools accessible to experimenters in agriculture, biology and other area of application. But perhaps Snedecor's most remarkable contribution was his book co-authored with Cochran, Statistical Methods Applied to Experiments in Agriculture and Biology, published first in 1937, and arguably the most popular and influential statistics text ever written.

Sir Ronald Aylmer Fisher (1890-1962): An English Statistician, biometrician and geneticist. Arguably, he was the greatest statistician of the twentieth century. Fisher transformed the Statistics of his day from modest collection of useful ad hoc techniques into a powerful and systematic body of theoretical concepts and practical methods. He is well known for his contributions to statistics by creating Fisher's exact test, the F-distribution and method of maximum likelihood. In 1925, the first edition of his Statistical methods for Research Workers appeared. Ander Hald called him "a genius who almost single-handedly created the foundations for modern statistical science".

Frank Wilcoxon: (1892-1956): He was an American statistician. He was initially a chemist and self-taught statistician. The first edition of statistical method for Research workers by Sir R.A further aroused his statistical interest in 1925. His statistical work concentrated on devising methods of testing, were simples and easy to understand (the hall marks of
nonparametric tests). Wilcoxon is remembered now in the context of the Wilcoxon signed-ran tests introduced in a 1945 paper.

Egon Sharpe Pearson (1895-1980): He was an English mathematical statistician, the only son of Karl Pearson. His association with Jerzy Neyman started in 1926 and led to the Neyman-Pearson lemma and the development of a standard approach to hypothesis testing. He also touched on problems of a quality control and operation research.

Gertude Mary Cox (1900-1978): She is an American and statistician. She became interested in statistics because her computing experience. Gertrude Cox is principally known as a gifted statistical administrator as well as for her contributions to psychological statistics and experimental design. At a time, she was put in charge of establishing a Computing Laboratory and consulted in experimental design. In 1934 she began to teach design of experiment a course that would become renowned. She later published her design material in a collaborative effort with William Cochran. She was a pioneer in the use of computer programs, with her staff developing many of the early SAS algorithms. She was joint author with William Cochran, of the statistical classic Experimental Designs.

Abraham Wald (1902-1950): He is Hungarian geometer and statistician. He turned his attention to statistical decision theory and made important advances in the theory of sequential sampling. His work is now especially remembered by statisticians in the context of the Wald-Wolfowitz run test; Wald’s equation and Wald’s distribution (also called the inverse normal distribution). In memory of Abraham Wald, the Wald Lectureship is awarded annually by the Institute of Mathematical Statistics (IMS).

Andrei Nikolaevich Kolmogorov (1903-1987): He is a Russian mathematician, statistician and probabilist. He obtained results in Fourier series (1922), logic (1925) and probability (1929). He also worked on such practical projects as turbulence, the motion of the Planets, telecommunications scheduling, among others. His work is now especially remembered by statisticians in the context of the Chapman-Kolmogorov equation for Markov processes and the nonparametric kolmogorov-Smirnov test for a specified distribution.

Rensis Likert (1903-1981): He is an American psychologist and statistician. His work involved the interpretation of surveys and the handling of questionnaires, he developed reliable and efficient multistage probability area sampling methods for households and individuals; used with high standards for many national surveys of behaviors and attitudes. Likert is now especially remembered by statisticians in the context of the Likert Scale. A scale, usually of approval or agreement, used in questionnaire.
Rensis Likert (1903-1981): He is an American psychologist and statistician. His work involved the interpretation of surveys and the handling of questionnaires, he developed reliable and efficient multistage probability area sampling methods for households and individuals; used with high standards for many national surveys of behaviors and attitudes. Likert is now especially remembered by statisticians in the context of the Likert Scale. A scale, usually of approval or agreement, used in questionnaire.

Andrei Nikolaevich Kolmogorov (1903-1987): He is a Russian mathematician, statistician and probabilist. He obtained results in Fourier series (1922), logic (1925) and probability (1929). He also worked on such practical projects as turbulence, the motion of the Planets, telecommunications scheduling, among others. His work is now especially remembered by statisticians in the context of the Chapman-Kolmogorov equation for Markov processes and the nonparametric kolmogorov-Smirnov test for a specified distribution.

Rensis Likert (1903-1981): He is an American psychologist and statistician. His work involved the interpretation of surveys and the handling of questionnaires, he developed reliable and efficient multistage probability area sampling methods for households and individuals; used with high standards for many national surveys of behaviors and attitudes. Likert is now especially remembered by statisticians in the context of the Likert Scale. A scale, usually of approval or agreement, used in questionnaire.

Maurice Stephenson Bartlett (1910-2002): He was an English statistician who made particular contributions to the analysis of spatial data. In 1933, Bartlett was recruited by Egon Pearson to the new Statistics Department at University College, London. Pearson was already working with Jerzy and Ronald A. Fisher. He was stimulated by all of them, especially by the work of Fisher. In 1934, Bartlett became Statistician at the ICI Agricultural Research Station at Jealott’s Hill. He remembered ICI as the best working environment of his career where he published a book on Stochastic Processes. He is also known for his work in the theory of statistical inference, time-series analysis and stochastic process.

Milton Friedman (1912-1937): He was an American economist, statistician and academic. He is best known for his theoretical and experimental research. He is popularly remembered for the Friedman’s two-way ANOVA by ranks which he proposed in 1937.

Lee Joseph Gonbach (1916-2001): He was an American psychometrician. His research included work on statistics evaluation and instruction. He produced many of his influential work in 1948. Prominent among these is the “Alpha” i.e. measure of reliability.

DATA COLLECTION, ANALYZING, INTERPRETATION AND MAKING CONCLUSION.

Data Collection

Data Collection Procedures

The two ways of Data collection procedure are:

i. Census: is a procedure of collecting data in which every unit of the population is considered. All statistical computations that are based on the entire population are called population parameters.

For Example, the population mean ($\mu$), population variance ($\sigma^2$) population correlation coefficient ($\rho$), among others.

ii Sample survey: Is a data collection procedure, where the data are obtained from the population. In any case, the sample survey is aimed at estimating the population. Any computation based on the sample is called a sample statistics are simple mean ($\bar{X}$), sample variance ($s^2$), sample correlation coefficient ($\gamma$), among others.

Sources of Statistical Data

Data may be gathered from two main sources namely.

i. Primary data: They are those data gathered directly from the source that is when data is collected directly from the observed values. For example, data collected through questionnaires, interviews, observation or experiments.

ii. Secondary data: Secondary data they are data collected from other sources, but essentially not directly from observed values.

For example data collected through transcription for documentation from existing records. Data collected from Magazines, Journals, library, and archives, among others.

Method of Data Collection

Data are generally needed, collected and analyzed to provide useful and meaningful information for planning, making decision and policy implementation. The entire planning and execution of a survey or experiment depends on the type of data needed which is greatly influenced by the method through which the data are collected. The decision and choice of method of data collection should be arrived at after considering the aims and objectives.
of the survey or experiment, the nature of the information needed the population under study, the degree of accuracy desired, practicability of results, time and cost. The methods are as follows:

i. Mail/postal questionnaire method involved the preparation of questionnaires on the subject matter of the survey.

ii. Personal interview method, required the interviewer to ask prepared set of questions from the respondent and record the answer.

iii. Telephone interview method; involves the use of modern telephone system, an interviewer can obtain information from respondents by making use of a telephone.

iv. Direct observation method; is the collection of data by physical observation or measurement of the units, items or respondents under study.

v. Experimental method; the desired data in this method are collected through the design and analysis of experiment.

vi. Transcription or Documentation method; this method is used when the data is needed for a specified survey.

Data Editing
The concept of data editing refers to some deciphering processes after data collection with a view to detecting errors, omissions and inconsistence, the work of editing requires skills and impartiality of high degree, the four purpose of editing are as stated below:

i. Editing for consistency

ii. Editing for accuracy

iii. Editing for uniformity

iv. Editing for completeness

With particular reference to primary data, the returned questionnaire filled by the respondents is scrutinized at the early stage so as to sistencies. In editing secondary data, the data analysts must bear in mind that they are using someone else’s work.
The y are actually data collected by somebody else; not directly by the data analysts. Hence, there is a limit to the degree of reliability one can place on secondary data. Because of this, it is necessary to scrutinize every secondary data in the light of the following points:

i. The purpose for which the data are originally collected to serve and how the data were used earlier

ii. The types and purpose of the institution that publishes the data as a routine.
iii. The nature of the data themselves, whether the data are based or not, simple or complex.
iv. The unit of measurement in which the data are expressed.
v. The degree of accuracy of the data.
vi. Relating the data to the problem under study.
vii. Do the data refer to homogenous omission or not.

Sources of Errors in Data
The two major sources of errors in data collection are:
i. Sampling Errors: They are variations due to the fact that only n units are measured instead of the entire N units of the population.
ii. Non-sampling Errors: They are errors due to lack of correct and adequate data.

Data Analyzing
In analyzing some of the statistical data, some of the analytical method may be considered such as:
1. Measure of central tendency: involves the Arithmetic mean, median, mode, Geometric Harmonic mean; they are values somewhere midway between the highest and the lowest observation.
2. Measure of Dispersion: are meant to measure the degree to which numerical values are spread around an average such as: Range, mean deviation, variance standard deviation, semi interquartile range and coefficient of variation.
3. Graphical presentation of data: is the simplest method of displaying statistical information on graphs, charts and a private a d to statistical analysis. The data may be presented in any of pie chart, Bar chart, Histogram, Frequency polygon, and cumulative curve (ogine)
4. Probability distribution: such as Binomial, poison, Normal, Gamma, Beta, Geometric, and Hyper geometric among others are used to measure the degree of tendency in the occurrence of an event.

Note:
i. Skewness And Kurtosis: they both give the sense of the general from or shape of a frequency distribution, skewness is the degree of departure while kurtosis (convexity of a curve) is the degree of Flareness of a distribution, usually both are relative to a normal distribution.
ii. Some Continuous Distribution: there are some very important continuous distributions which play significant role in the field of Statistics. Each of these distributions has a specialized probability density function with some
peculiar properties. Some of these specialized continuous distributions have their values tabulated for easy computations; such as student's t-Distribution, chi-square Distribution, Hoelling's T^2 Distribution, Bartlett test, Cochran C-test, Fisher's F-max test and F-Distribution.

iii. Goodness of fit tests: is one of the important applications of the chi-square test is that of testing the goodness of fit of a set of data to a specific probability distribution. In testing the goodness of fit of a set of data, the actual frequencies that would be theoretically expected to occur if the data followed a specific probability distribution.

5. Statistical Sampling: Sampling theory is the study of relationships existing between a population and samples drawn from the population. It is useful in estimation of unknown population quantities (such as population mean and variance) often called population parameters from knowledge of corresponding sample quantities (such as sample mean & variance) often called sample statistics or statistics.

Modes of sampling are:

i. Sampling with replacement (SWR): it is a sampling techniques where an element may occur more than once.

ii. Sampling without replacement (SWOR): it is a sampling techniques where each of the elements in the sample occurs only once.

6. Test Hypothesis: a statistical hypothesis; is an assertion, which may or may not be true concerning one or more population.

i. Null and Alternative Hypothesis: The hypothesis formulated somewhat with the hope of rejecting is called null hypothesis denoted by H₀ leads to the acceptance of an alternative hypothesis denoted by H₁.

ii. Type I and type II Errors: They are known as the Hypothesis errors. Type I error is committed if null hypothesis H₀ is rejected, when it is actually true. While type II error is committed if H₀ is accepted when it is actually false.

7. Correlation: it is the measure of the degree and direction of linear relationship existing between two or more variable capable of quantitative measurement. The strength of a relationship or the association, between two variables is typically measured by the coefficient of correlation, whose value range from -1 for a perfect negative correlation up to +1 for a perfect positive correlation.

However, person's product-moment correlation coefficient, spear man's rank correlation coefficient, & kendall's Rank correlation coefficient are used between two set say X and Y; while canonical correlations are used to measure the interrelationship between two sets of data say; dependent sets (X₁) and independent sets (Y₁).
8. **Regression Analysis:** it is a statistical methodology that utilizes the relationship between two or more quantitative variables so that one variable can be predicted from the other variable or variables. Regression analysis attempts to establish a relationship between a numerical variable (Independent/predictor variables). The relationship is pressured to be of a known functional the type (e.g. linear, quadratic, etc) and the corresponding function is called the regression function or regression equation. The functions parameters are chosen to minimize the "prediction error".

i. **Simple Regression:** it involves models with only one dependent variable and one independent variable and the regression function is linear.

ii. **Multiple Regression:** Multiple regression is not just one technique but 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 (usually continuous). Multiple regression is based on correlation, but allows a more sophisticated exploration of the interrelationship among a set of variables. This makes it ideal for the investigation of more complex real-life, rather than laboratory-based, research questions. Multiple regression is not used as a fishing expedition. It involves models with only one dependent variable and two or more independent variable. The objective in a multiple regression problem is essentially the same as for a simple regression.

iii. **Logistic regression:** Logistic regression allows testing models in predicting categorical outcomes with two or more categories. The predictor (independent) variables can be either categorical or continuous, or a mix of both in the one model.

9. **Analysis of Variance (ANOVA):** Analysis of variance is so called because it compares the variance (variability in scores) between the different groups (believed to be due to the independent variable) with the variability within each of the groups (believed to be due to chance). An F ratio is calculated, which represents the variance between the groups, divided by the variance within the groups. A large F ratio indicates that there is more variability between the groups (caused by the independent variable) than there is within each group (referred to as the error term). A significant F test indicates that the null hypothesis can be rejected, which states that the population means are equal. It does not, however, tell which of the groups differ. It is usually used for test of significance for regression parameters. It analyzes the variation in the set of data say Y into its components parts; one part due to relationship with X and the other parts due to error.

10. **Nonparametric statistics:** is a statistical procedure that has certain desirable properties that hold under relatively mild assumptions regarding the underlying populations from which the data are obtained. These
procedures make less of many assumptions which are made under parametric statistical tests. Some of the nonparametric are: Mann-Whitney test, Friedman test, Kruskal-Wallis test and Theil’s method for linear regression among others.

11. Time series Analysis: Time series data, as the name suggests, are data that have been collected over a period of time on one or more variables. Time series data have associated with them a particular frequency of observation or collection of data points. The frequency is simply a measure of the interval over, or the regularity with which, the data are collected or recorded. Some of the model used in solving the time series are MA(p) AR(q), ARMA(p,q), ARIMA (p,l,q), among others.

SOME STATISTICAL PACKAGES FOR DATA ANALYSIS.
Statistical packages usually help statisticians to use a wide range of statistical techniques by automating the data processing and computational borders of such techniques. Most of the methods of analysis may be carried out using standard statistical packages or languages. The widely available packages include.

(1). SPSS (Statistical Package for the Social Sciences) (2). SAS (Statistical Analysis System)
(3). GENSTAT (General Statistical Package) (4). BMDP
(Biomedical Data Packages)
(5). NCSS (Number Cruncher) Statistical System) (6). MINITAB
(7). STATISTIX
(8). SYSTAT (9). STATISTICA (10). R-Statistical Software (11). S-
Plus (12). Stat Direct
(13). LISREL (Linear Structural Relationship) system. (14). E-Views
(15). J-MULTI
(16). AMDS (Advanced Moment Structures) (17). EXCEL.

Many of these packages, such as SAS and SPSS offer facilities for data management tasks such as editing and checking. Another package S-Plus BMDP provide data interactive data analysis system. Together with a programming language S. Start Direct package covering many of the methods for medical applications.

In addition, the major statistical computing packages which cover many of the standard methods, there are many other packages or programs suitable for some of the more specialized tasks. Occasional references to these packages are made throughout the course, e.g. Eviews and J-multi for time series analysis, among others.
Although computers are increasingly used for data analysis, with smaller sets of data, it is often convenient to use a calculator, the most convenient form of which is the pocket scientific calculators, that perform a high speed of all the basic arithmetic operation including some range of mathematical functions such as square, square roots, exponential, logarithms etc. An additional feature particularly useful in statistical work is the automatic calculation and accumulation of sums of squares of numbers, some calculators have a special range of extended facilities for statistical analysis.

**DATA INTERPRETATION**

Interpretation of data is one of cardinal emphasis required after analysis, while the first irreplaceable step in interpreting data is to calculate basic descriptive statistics for the primary variables under investigations and to develop a visual depiction of the relations among variables. These descriptive statistics and graphic visual depiction serve as two prerequisites towards a comprehensive inferential analysis of the data.

The growth of statistics especially in the method of data interpretation during the past sixty years, started in the field of biological research, but the nature of the problems encountered has caused the new methods to be applied to science, medical, psychological, economic data, and to some extent in physics, chemistry and engineering.

The biologist in his routine work is confronted with the difficulty that the measurements after identical treatment of two animals or plants, apparently similar in all aspects, can give widely different results.

For instance, an insecticide applied to two batches of insects can give different percentage kills, while the change in crop yield due to the application of fertilizer may vary widely from field to field. The natural variability more or less masks any real effect, and makes the drawing of conclusion from collected data difficult.

For instance, also the physicist can frequently use a simple observation as a basic for further work, the biologist can seldom rely upon one observation. Thus it was in the biological field that the laws of variability and their application to the interpretation of data were initially investigated, but the general nature of these laws has made the new statistical techniques of much wider use.

**Drawing Conclusion**

Statistics is the most veritable tool for planning, policy formation and decision making. In everyday life, in every sector, decisions are taken and plans are made, but the success of every plan depends on the accuracy, analysis and
interpretation of the available data with the respective statistical methodologies involved in drawing a conclusion. This shows that statistics plays a vital role in all findings, planning and decision making processes.

Summary and Conclusion
The history of statistics suggests two important lessons. First, methodological research flourished in the hands of those of who were not only highly focused on problems arising from real data, but who were themselves immersed in generating highly valuable data. Second, although theoretical statistics research can be natural in isolating only an applied orientation has made it possible for such efforts to reach new heights.

The method involved in the collection of data as practiced by the past scholars tends to avoid commission of errors in all regards; the careful use of standard statistical methods enables to accurately describe the findings of scientific, technology or socio-economic research. Also, the use of some statistical software would aid accurate outcome, interpretation, and making a substantial decision and planning. Some people view statistics as a field of hocus-pocus whereby a person in the known overwhelm the armature. The word statistics has different meaning to people of diverse background and interest. To some people it is a way of collecting and displaying large amount of numerical information. While to some other group it is a way of making decision in the field of uncertainty.

However, Statistics has been defined as the science which deals with the collection, collation or organization, summarization, presentation, analysis and interpretation of numerical data with the aim of drawing valid and logical conclusions and making reasonable decisions on the basis of the analysis.

References


SPSS Survival Manual Original 2007 version editable