

1. Information, which are not numeric in nature are called qualitative variable
2. This assignment of numerical values to information is called coding
3. A variable is any characteristic of an object or concept that is capable different values or falling into more than one distinct category.
4. any characteristic of an vary in time and space is called a variable.
5. Quantitative variable can be classified into two major types
6. Discrete variables are variable those values are whole numbers or integers.
7. Continuous variables are variables that assume any value ii'ithin an interval or r have the property of infinite divisibility.
8. There are four measurement scales available as instruments for measuring variable The scales are nominal, ordinal, interval and Nominal scale
9. **Nominal** : This scale groups the objects into distinct categories to facilitate referencing
10. **Ordinal**: This scale ranks or orders the mutually exdusive categories of the variables according to the importance attached to each category. This scale has all the properties of the nominal scale plus the additional property of ordering or ranking the categories
11. **Interval**: This has the combined properties of the nominal and ordinal scale plus the additional property of measuring the distance or interval between two measurements. This scale gives information on how much one category is more or less than the other. It has no absolute zero.
12. Interval variables are quantitative and may be discrete or continuous
13. **Ratio scale** This scale has all the properties of the nominal, ordinal and interval scales including the additional property of having an absolute zero point. This is the highest level of measurement.
14. The second most important characteristics which describe a set of data is the amount of variation, scatter, or spread in the data.
15. The variation or dispersion can be said to measure the degree of uniformity of observations in a given set of data.
16. The greater the variation, the more un-uniform the observations in a given set of data
17. The Range (R) of a given set of ungrouped data can be determined from an ordered array as the difference between the highest observation and the lowest observation in a distribution..
18. The range can be an unsatisfactory measure of dispersion because it is affected by extreme values or items which renders it unrepresentative of majority of the set of data.
19. Unlike the range, quartile deviation does not take extreme values or items. Quartiles are the boundaries separating the items in a given distribution or set of data into quarters.
20. There are, therefore, three quartiles: the *lower quartile* (at the 25 percent mark); the *median* (at the 50 percent mark); and, the *upper quartile* (at the 75 percent mark)
21. The quartile deviation referred to as the semi-interquartile range is defined as one-half the difference between the upper quartile and the lower quartile
22. the standard deviation is the most useful measure of variation. It can be defined as the square root of the variance for a given set of data.
23. the standard deviation is the most useful measure of variation. It can be defined as the square root of the variance for a given set of data.
24. Unlike other measures of variability, the coefficient of variation is a relative measure. It is particularly useful when comparing the variability of two or more sets of data that are expressed in different units of measurements. The coefficient of variation measures the standard deviation relative to the mean and is computed by:
25. Coefficient of Variation = $CV = S \times 100\%$

26. The coefficient of variation is also useful in the comparison of two or more sets of data which are measured in the same units but differ to such an extent that a direct comparison of the respective standard deviations is not very helpful
27. The measures of skewness are generally called Pearson's first coefficient of skewness and Pearson's second coefficient of skewness.
28. Measures of skewness are used in determining the degree of asymmetry of a distribution; a distribution which is not symmetrical is said to be skewed.

$$Sk = \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$$

29. Kurtosis measures the degree of peakedness of a distribution. It is usually taken relative to a normal distribution.
30. There are usually three types of kurtosis namely. LEPTOKURTIC, PLATYKURTIC and MESOKURTIC. \
31. The mesokurtic is otherwise known as normal distribution curve i.e. the curve that is moderately distributed.
32. The moment coefficient of kurtosis is used to calculate the peakedness of a distribution
33. Decision analysis is the modern approach to decision making both in economics and in business.
34. Decision analysis : It can be defined as the logical and quantitative analysis of all the factors influencing a decision.
35. The art of problem solving and decision making is base on common sense.
36. There are two (2) main ways of approaching problems and obtaining solutions. 1. Analytical Thinking 2. Creative Thinking.
37. ANALYTICAL THINKING: Seeks to improve on a given situation
38. CREATIVE THINKING: This considers end rather than means
- 39.
40. **Critical Examinations: Is the logical approach, it answers questions like What, Who, Where, How, When and Why.**
41. **Brain Storming:** This is a method base on two head is better than one.
42. Brain Storming involves conference techniques by which a group of people attempts to find solution for specified problems by amazing all ideas spontaneously contributed by its members. It is a free thinking meeting.
43. Steps on Brain Storming are as follows: Orientation, Consideration, Speculation (opinion), Recommendation
44. **Analogies: This is the comparison of one thing with another that has similar features**
Types of analogy
45. Personal Analogy: Is putting self in place of object.
46. Direct Analogy: This to compare things in nature with the situation and use nature's solution to provide a lead to a suitable solution.
47. Symbolic Analogy: Uses objects and impersonal images.
48. Fantancy Analogy: The problem is transformed into the realism of fantancy.
49. **Morphological Approach or Attribute Listening: This method looks for the attributes or qualities of the product i.e. comparison of the best one.**
50. The alternative actions and their corresponding payoffs are usually known to the decision-maker in advance.
51. Economic variables are referred to, in decision analysis, as *states of nature* as they represent different events that may occur, over which the decision maker has no control.
52. When the state if nature, si, whether known or unknown, has no influence on the outcomes of given alternatives, we say that the decision maker is operating under **certainty**. Otherwise, he/she is operating under uncertainty
53. Decision making *under certainty* appears to be simpler than that under uncertainty.
54. Under certainty, the decision maker simply appraises the outcome of each alternative and selects the one that best meets his/her objective.
55. If the number of alternatives is very high however, even in the absence of uncertainty, the best alternative may be difficult to identify.
56. Decision making *under uncertainty* is always complicated. It is the probability theory and mathematical expectations that offer tools for establishing logical procedures for selecting the best decision alternatives.
57. Though statistics provides the structure for reaching the decision, the decision maker has to inject his/her intuition and knowledge of the problem into the decision-making framework to arrive at the decision that is both theoretically justifiable and intuitively appealing.
58. A good theoretical framework and commonsense approach are both essential ingredients for decision making under uncertainty.
59. Associated with each action is a list of payoffs. If an action does not involve risk, the payoff will be the same no matter which state of nature occurs.

60. The payoffs associated with each possible outcome in a decision problem should be listed in a *payoff table*, defined as a listing, in tabular form, of the value payoffs associated with all possible actions under every state of nature in a decision problem
61. The payoff table is usually displayed in grid form, with the states of nature indicated in the columns and the actions in the rows
62. A decision-making procedure, which employs both the payoff table and prior probabilities associated with the states of nature to arrive at a decision is referred to as the Expected Monetary Value decision procedure
63. By *prior probability* we mean probabilities representing the chances of occurrence of the identifiable states of nature in a decision problem prior to gathering any sample information.
64. The *expected monetary value decision* refers to the selection of available action based on either the expected opportunity loss or the expected profit of the action.
65. Decision makers are generally interested in the optimal monetary value decisions
66. The optimal expected monetary value decision involves the selection of the action associated with the minimum *expected opportunity loss* or the action associated with the maximum *expected profit*, depending on the objective of the decision maker
67. The concept of expected monetary value applies mathematical expectation, where opportunity loss or profit is the random variable and the prior probabilities represent the probability distribution associated with the random variable.
68. Index numbers are indicators which reflect the relative changes in the level of certain phenomenon in any given period (or over a specified period of time) called the current period with respect to its value in some fixed period called the base period selected for comparison.
69. Index number may be classified in terms of the variables they measure. They are generally classified into three categories:
70. 1. **Price Index Number**: The most common index numbers are the price index numbers which study changes in price level of commodities over a period of time. They are of two types:
71. (a) Wholesale price index number – They depict changes in the general price level of the economy.
72. (b) Retail Price Index Number – They reflect changes in the retail prices of different commodities. They are normally constructed for different classes of consumers.
73. 2. **Quantity Index Number** – They reflect changes in the volume of goods produced or consumed
74. 3. **Value Index Number** – They study changes in the total value (price X quantity) e.g. index number of profit or sales.
75. The base period is the previous period with which comparison of some later period is made.
76. The index of the base period is taken to be 100.
77. Simple (unweighted) Aggregate Method – Aggregate of prices (of all the selected commodities) in the current year as a percentage of the aggregate of prices in the base year.
78. Weighted Aggregate Method - In this method, appropriate weights are assigned to various commodities to reflect their relative importance in the group.
79. Dorbish-Bowley Price Index – This index is given by the arithmetic mean of Laspeyres's and Paasche's price index numbers.
80. Fisher's Index is termed as an ideal index since it satisfies time reversal and factor reversal test for the consistency of index numbers
81. Primary data – These are data collected directly from the field of enquiries by the user(s) themselves
82. Secondary Data: These are data which have been collected by someone else or some organization either in published or unpublished forms
83. Cross-sectional data – These are data collected for cross-section of subjects (population under study) at a time.
84. Time-series data – These are data collected on a particular variable or set of variables over time e.g. a set of Nigeria's Gross Domestic Product (GDP) values from 1970 to 2012.
85. Researchers collect data in order to test hypotheses and to provide empirical support for explanations and predictions
86. Generalizations are important not only for testing hypotheses but also for descriptive purposes.
87. Typically, generalizations are not based on data collected from all the observations, all the respondents, or the events that are defined by the research problem as this is always not possible or where possible too expensive to undertake
88. The entire set of relevant units of analysis, or data is called the population.
89. When the data serving as the basis for generalizations is comprised of a subset of the population, that subset is called a sample.
90. A particular value of the population, such as the mean income or the level of formal education, is called a parameter; its counterpart in the sample is termed the statistic.
91. The major objective of **sampling theory** is to provide accurate estimates of unknown values of the parameters from sample statistics that can be easily calculated.
92. a population is the “aggregate of all cases that conform to some designated set of specifications”

93. The specific nature of the population depends on the research problem
94. A single member of a sampling population (e.g a household) is referred to as a sampling unit.
95. Once researchers have defined the population, they draw a sample that adequately represents that population.
96. Ideally, the sampling frame should include all the sampling units in the population.
97. The accuracy of a sample depends, first and foremost, on the sampling frame.
98. **Sample Design** The essential requirement of any sample is that it be as representative as possible of the population from which it is drawn
99. In non-probability sampling, there is no assurance that every unit has some chance of being included.
100. Three major designs utilising non-probability samples have been employed by social scientists: convenience samples, purposive samples, and quota samples.
101. **Convenience samples:** Researchers obtain a convenience sample by selecting whatever sampling units are conveniently available
102. With purposive samples (occasionally referred to as judgement samples), researchers select sampling units subjectively in an attempt to obtain a sample that appears to be representative of the population
103. **Quota samples: The chief aim of quota sample is to select a sample that is as similar as possible to the sampling population**
104. Four common designs of probability samples are simple random sampling, systematic sampling, stratified sampling, and cluster sampling.
105. Simple random sampling – is the basic probability sampling design, and it is incorporated into all the more elaborate probability sampling designs. Simple random sampling is a procedure that gives each of the total sampling units of the population an equal and known nonzero probability of being selected.
106. Systematic Sampling: It consists of selecting every kth sampling unit of the population after the first sampling unit is selected at random from the total of sampling units
107. Systematic sampling is more convenient than simple random sampling. Systematic samples are also more amenable for use with very large populations or when large samples are to be selected.
108. Stratified Sampling: Researchers use this method primarily to ensure that different groups of population are adequately represented in the sample
109. stratified sampling considerably reduces the cost of execution. The underlying idea in stratified sampling is to use available information on the population “to divide it into groups such that the elements within each group are more alike than are the elements in the population as a whole.
110. **Cluster sampling:** it is frequently used in large-scale studies because it is the least expensive sample design.
111. Cluster sampling involves first selecting large groupings, called clusters, and then selecting the sampling units from the clusters.
112. The clusters are selected by a simple random sample or a stratified sample.
113. A sample is any subset of sampling units from a population. A subset is any combination of sampling units that does not include the entire set of sampling units that has been defined as the population. A sample may include only one sampling unit, or any number in between
114. **Standard error** Some people called it *error margin* or *sampling error*.
115. The concept of standard error is central to sampling theory and to determining the size of a sample. It is one of the statistical measures that indicate how closely the sample results reflect the true value of a parameter.
116. There are three methods of data collection with survey and these are mail questionnaires, personal interviews, and telephone interviews
117. Mail questionnaire: It is an impersonal survey method. Here, survey instrument (the questionnaire) is mailed to the selected respondents and the questionnaires are mailed back to the researcher after the respondents must have filled it up.
118. The personal interview is a face-to-face, interpersonal role situation in which an interviewer asks respondents question designed to elicits answers pertinent to the research hypotheses.
119. **Telephone interview** It is also called telephone survey, and can be characterised as a semi-personal method of collecting information.
120. In comparison, the telephone is convenient, and it produces a very significant cost saving.
121. Point Estimation when a single numerical value of the statistic is used as an estimate of the exact population value, we have a point or target estimate.
122. An estimate is value of the sample statistic which is taken as an approximation of the parameter value.
123. An estimator refers to the formula or statistic which has been chosen to provide an estimate of the population value.
124. The mean, mode, medium, variance etc are examples of point estimates.

125. Correlation can be defined as the branches of statistics that deals with mutual dependence or inter-relationship of two or more variables.
126. If the value of two variables such that when one changes, the other changes too, then the variable are said to be correlated.
127. correlation implies that variation in one variable, when there is a variation in other variable
128. The degree of relationship existing between two variables is called simple correlation. While the degree of relationship that connected three or more variables together is called Multiple correlation
129. **Perfect Positive Correlation** This can be defined as the situation where all the scatter points passes through a straight line none of the points deviated from the normal curve and positive slope.
130. **Perfect Negative Correlation:** This indicates that all the points passes through the normal straight line and non deviated from the line. The curve shown downward slope of units.
131. **Strong Positive Correlation:** In these case, most of the scatter points passes through the straight line, although there are few deviation from the straight line, but the deviation are very close to each other.
132. **Strong Negative Correlation:** In a strong negative correlation, some of the points passes through the straight line and all other scatter point are very close to the straight line, it has a negative slope which is very close to unity
133. **Weak positive correlation:** In these case the points are deviated from each other so that each of the scatter points are for the depart from each other and the association is weak.
134. **Weak negative correlation:** In a weak negative correlation, there are serious deviations of scatter points and the points slope downward. It has a negative slope and not close to unity.
135. **No Correlation:** The scatter point at random and did not form any regular pattern for recognition by any straight line. There is no association between the variables.
136. Rank correlation is used to determine the extent at which the variable are correlated. This idea was employed by Spearman's rank correlation coefficient, which is computed by using this formula.
137. Spearman's coefficient of correlation assumes the data to be at least interval scale. Chalse – Spearman, a British statistician, introduced a measure of correlation for ordinal- level data known as Spearman's rank-order correlation coefficient (i.e. A measure of relationship between two sets of ranked data)
138. Regression analysis can be defined as the relationship between two or more variables. This relationship has to do with the changes that result from a change in one of the related variables
139. Basically, there are two types of regression: i. Simple (linear) regression ii. Multiple (non linear) regression
140. **Simple (linear) regression** This involve only two variables and the relationship between them tends towards a fixed direction
141. **Multiple (non linear) regression** This also involved more than two variables in the regression model or equation.
142. 1. Freehand method In this method, regression line is fitted into the scatter diagram
143. The scatter diagram is the graphically representation of relationship which exists between two variables by drawing a lines of best fit through the various points which are estimate from the relationship x and y.
144. Least Square Method: This is the mathematical method of determined the points estimate of 'a' and 'b' from the available sample points. This method is the most reliable of all the methods.
145. There are 3 main types of curve
146. 1. Exponential (Growth) curve: This is a situation whereby when a data is expected to grow by some proportion or percentage in each period
147. Hyperbolic Model (curve): This has a formular
148. Power curve mor model: This power model have the form of $y = ax^b$. Otherwise known as logarithms functions
149. the degree of relationship that connect three or more variables together are called multiple correlation regression.
150. A hypothesis can be defined as a conjectural statement a postulate, or a proposition about an assumed relationship between two or more variables.
151. Hypothesis testing or testing a hypothesis are used interchangeably. Hypothesis testing starts with a statement about population parameters such as mean.
152. Such assumption which is subject to testing either may be true or may not be true is called statistical hypothesis
153. The square of a standard normal variable is called a Chi-square variate with 1 degree of freedom, abbreviated as d.f.
154. The chi-square test can be used to determine how well theoretical distributions 9such as the normal and binomial distributions) fit empirical distributions (i.e. those obtained from sample data).

155. Karl Pearson in 1900, developed a test for testing the significance of the discrepancy between experimental values and the theoretical values obtained under some theory or hypothesis.
156. The term **Analysis of Variance** was introduced by Prof. R.A Fisher in 1920s to deal with problems in the analysis of agronomical data. Variation is inherent in nature. The total variation in any set of numerical data is due to a number of causes which may be classified as: (i) Assignable causes and (ii) chance causes
157. The variation due to assignable causes can be detected and measured whereas the variation due to chances is beyond the control of human and cannot be traced separately.
158. ANOVA test is based on the test statistic F (or variance ratio).
159. ANOVA as a tool has different dimensions and complexities. ANOVA can be (a) One-way classification or (b) two-way classification
160. Forecasts are based on past performances. In other words, future values are predicted from past values
161. 1. Short-term Forecasts. These are forecasts concerning the near future. They, are characterized by few uncertainties and therefore more accurate than distant future forecasts
162. 2. Long – term forecasts. These concern the distant future. They are characterized by more uncertainties than short – term forecasts.
163. 3. Extrapolation. These are forecasts based solely on past and present values of the variable to be forecast. In this case, future values are extrapolated from past and present values.
164. 4. Forecasts Based on Established Relationships between the variable to be forecast and other variables.

METHODS OF FORECASTING

165. There are generally used methods of forecasting: I. Moving Averages II. Trend lines or least squares.
166. Moving averages can be used to generate the general picture (or trend) behind a set of data or time series. The general pattern generated can be used to forecast future values.
167. **Time series** is a name given to numerical data described over a uniform set of time points. Time series occur naturally in all spheres of business activity.
168. The least – squares method provides a sound mathematical basis for choosing the best trend line; of all possible trend lines for a given set of time series.
- 169.