BIO206 - BIOSTATISTICS SUMMARY

_____ is a familiar and accepted part of modern world that is concerned with obtaining an insight into the real world by means of the analysis of numerical relationships
Statistics

Statistics is a familiar and accepted part of modern world that is _____
concerned with obtaining an insight into the real world by means of the analysis of numerical relationships.

The word statistics is used in _____senses.
Two

_____ refers to collections of quantitative information, and to methods of handling that sort of data i.e. descriptive statistics.
Statistics

_____ refers to the drawing of inferences about large groups on the basis of observations made on smaller one i.e. inferential statistics.
Statistics

_____ deals with ways of collecting, organizing, summarizing and describing quantifiable data, and methods of drawing inferences and generalizing upon them.
Statistics

_____ is used when the data that are being analysed using statistical tools, are derived from the fields of biological sciences: Medicine, Pharmacy, Biochemistry, Microbiology, Agricultural Sciences and other biology-related areas.
Biostatistics

_____ helps to explain this natural variation inherent in these fields of natural sciences.
Biostatistics

Statistics is concerned with the analysis of _____relationships
Numerical

Statistics is useful in almost all fields of human endeavour
True

The following are fields of life in which statistics is used

- sports
- public health
- education
- surveys
- operations research
- quality control
- estimation and prediction

The word statistics is used in _____senses
Two

_____ statistics refers to collections of quantitative information and methods of handling that sort of data
Descriptive

_____ statistics refers to the drawing of inferences about large groups on the basis of observations made on smaller one.
Inferential

Biostatistics is used when the data that are being analysed using statistical tools are derived from the fields of biological sciences:

- Medicine
- Pharmacy
- Biochemistry
- Microbiology
- Agricultural Sciences

_____ is regarded as a fundamental feature in natural sciences of biology, agriculture and medicine
Variation

_____is an important aspect of biostatistics that describe on how to collect, organize, summarize and analyze data
Experimental design

The aspect of biostatistics that describe on how to collect, organize, summarize and analyze data is called_____
Experimental design

To gain knowledge about secondly haphazard events, _____collect information for variables which describe the event
Statistician

A person who collects information for variables which describe an event for statistical analysis is called a_____
Statistician

A _____is a characteristics attribute that can assume different value
Variable

Quantitative variables can be grouped into_____
two

A _____ is a subgroup of the population selected for study.
sample

A sample is_____
a subgroup of the population selected for study

The following are the importance of sample / sampling
1. It saves the researcher time and money.
2. It enables the researcher to get information that he or she might not be able to obtain otherwise.
3. It enables the researcher to get more detailed information about a particular subject.

_____ assists in avoiding various forms of conscious and unconscious bias
Randomness

A stratified sample is a sample obtained by dividing the population into subgroups called_____
strata

The advantage of stratified sampling is that_____
it increases precision because all types of groups are represented through stratification and a heterogeneous population is made into a homogenous one

A _____ is a sample obtained by selecting a preexisting or natural group
cluster sample

The following are the advantage of using cluster sampling
1. A cluster sample can reduce cost
2. It can simplify field work.

3. It is convenient.

_____ involves taking an item as a sample from a larger population at regular intervals.
Systematic or skip sampling

_____ type of sampling involves selecting a sample in proportion to the different groups in the population under study.
Proportionate sampling

_____ is the distribution of values from a mass of samples, one value per sample.
Sampling distribution

_____ or counting gives rise to raw data.
Measurements

Measurements or _____ gives rise to raw data.
Counting

Measurements or counting gives rise to_____.
raw data

The most convenient method of organizing data is to construct a _____.
frequency distribution

_____ is the number of occurrences of an element in a sample and is symbolized by f.
Frequency

A _____ is the organization of raw data in table form, using classes and frequencies.
frequency distribution

_____ is used for data that can be placed in specific categories, such as nominal or ordinal-level data.
Categorical frequency

The reasons for constructing a frequency distribution are:
1. To organize the data in a meaningful, intelligible way.
2. To enable the reader to determine the nature and shape of the distribution.
3. To facilitate computational procedures for measures of average and spread.
4. To enable the researcher to draw charts and graphs for the presentation of data.
5. To enable the reader to make comparisons among different data sets.

_____ is a branch of mathematics which as a general concept can be defined as the chance of an event occurring.
Probability

_____ is the basis of inferential statistics.
Probability

A _____ is a scatter of related values, such as the assortment of weights in a group of cattle.
distribution

A _____ shows us how many times given values in a range of values occur.
frequency distribution

A _____ is very similar because it shows us how probable given random variable values in a range of such values are.
Probability distribution

_____ is the most important distribution in statistics.
The normal distribution

The _____ is a particular pattern of variation of numbers around the mean.
normal distribution

_____ is concerned with results obtained by taking measurements on continuous random variable
Normal distribution

Variables can be classified into two broad categories: _____
qualitative and quantitative variables

_____ are variables that can be placed into distinct categories according to some characteristic or attribute
Qualitative variables

For example, if subjects are classified according to gender (i.e. male or female), then the variable gender is qualitative

_____ variables are numerical and can be ordered or ranked
Quantitative

In statistics, variables which are numerical and can be ordered or ranked are known as _____
Quantitative variables

Quantitative variables can be grouped into _____
two

Quantitative variables can be grouped into two: _____
Discrete and Continuous variables

_____ are variables that can be assigned values such as 0,1,2,3 (integers) and assume values that can be counted
Discrete variables

Examples of discrete variables include the following

- number of children in a family
- number of birds in a pen
- number of trees in a garden
- number of animals per litter

_____ are variables which assume all values between any specific values obtained by measurement
Continuous variables

These are examples of continuous variables

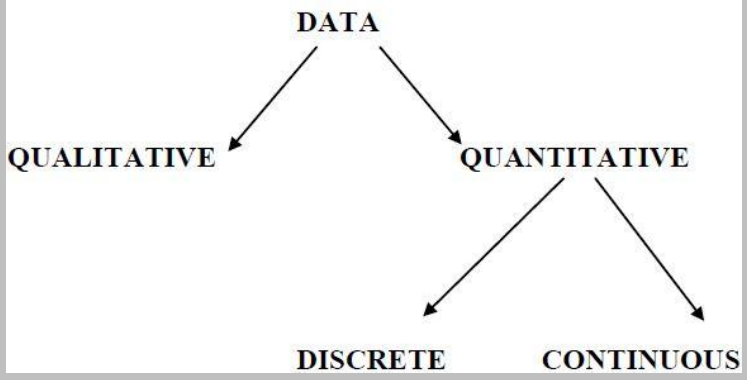
- Length
- Weight
- Height
- Yield
- Temperature
- Time

Time is an example of _____variables
continuous

_____ are the values (measurements and observations) that the variables can assume
Data

NounElearn www.nounelearn.com 07018870798 @NounElearn

: Summary of classification of variables.



The computation of a set of observations collected from a population mean (μ), population variance (σ2) and population standard deviation (σ) is called_____
Sampling

The parameters used to describe samples are the following

- mean (x)
- variance (s^2)
- standard deviation (s)

_____ is a random process of selecting a sample from a population selected for study
Sampling

_____is the computation of a set of observations collected from population mean, variance and standard deviation
Sampling

A random process of selecting a sample from a population selected for study is called_____
Sampling

A _____ is a subgroup of the population selected for study
sample

A subgroup of the population selected for study is known as a_____
sample

When a sample is chosen at random from a population, it is said to be an _____
unbiased sample

When a sample is chosen at random from a population, it is said to be an_____

unbiased sample

If a sample is selected incorrectly, it may be a _____sample
biased

A sample must be random in order to make valid inferences about the population
True

The following are importance of sample / sampling
1. It saves the researcher time and money
2. It enables the researcher to get information that he or she might not be able to obtain otherwise
3. It enables the researcher to get more detailed information about a particular subject

The following are types of sampling methods

- Systematic sampling
- Stratified sampling
- Clustered sampling
- Random sampling

For a sample to be a _____sample, every member of the population must have an equal chance of being selected
random

A sample in which every member of the population have an equal chance of being selected is called **Random sampling**

A _____is a sample obtained by dividing the population into subgroups, called strata
stratified sample

A _____is a sample obtained by dividing the population into subgroups according to homogenous (alike) characteristics and then selecting members from each stratum for the sample
stratified sample

Stratified sample is obtained by dividing the population into subgroups of homogenous characteristics called_____
strata

The advantage of stratified sampling is that_____
it increases precision

A cluster sample is a sample obtained by selecting a preexisting or natural group, called a_____
Cluster

A _____is a sample obtained by selecting a preexisting or natural group using the members in the cluster for the sample
cluster sample

These are the advantages of using a cluster sample over other types of sample:
1. A cluster sample can reduce cost
2. It can simplify field work
3. It is convenient

A sampling method which involves taking an item as a sample from a larger population at regular intervals is called _____
Systematic or Skip sampling

Sampling that involves selecting a sample in proportion to the different groups in the population under study is known as _____
Proportionate sampling

Systematic sampling is also called_____
skip sampling

In _____ sampling, successive units taken from production lines are sampled to ensure that products meet certain standards set by the manufacturing company
sequence

In_____, a very large population is given a questionnaire to determine those who meet the qualifications for a study
double sampling

A systematic sampling whereby the researcher uses a combination of sampling methods is referred to as_____
multistage sampling

The sampling distribution of the mean is determined by the nature of the _____and sample size
population

The sampling distribution of the mean is determined by the nature of the population and _____
sample size

The sampling distribution of the mean is determined by the nature of the _____
population and sample size

A _____ is the distribution of values from a mass of samples, one value per sample
sampling distribution

A sampling distribution is_____
the distribution of values from a mass of samples, one value per sample

When the sample size of a distribution is large then the sampling distribution of the mean approximates very closely to a_____
normal curve

_____is difficult to comprehend because it lacks organization, summarization, which renders it meaningless
Raw data

Raw data lacks organization and _____
summarization

Raw data lacks _____and summarization
organization

Raw data lacks_____
organization and summarization

A data which is meaningless because it lacks organization and summarization is called_____
raw data

The most convenient method of organizing data is to construct a_____
frequency distribution

_____ is the number of occurrences of an element in a sample
Frequency

Frequency is symbolized by letter_____
f

A _____is the organization of raw data in table form, using classes and frequencies
frequency distribution

The organization of raw data in tabular form using classes and frequencies is referred to as_____
frequency distribution

When data are collected in original form, that is as observed or recorded they are called_____
raw data

Data collected in original (or raw) form and observed or recorded are called_____
raw data

There are _____types of frequency distributions
two

Two types of frequency distributions that are most often used are the: _____
categorical and ungrouped frequency

The frequency distribution used for data that can be placed in specific categories, such as nominal or ordinal-level data is____
Categorical frequency

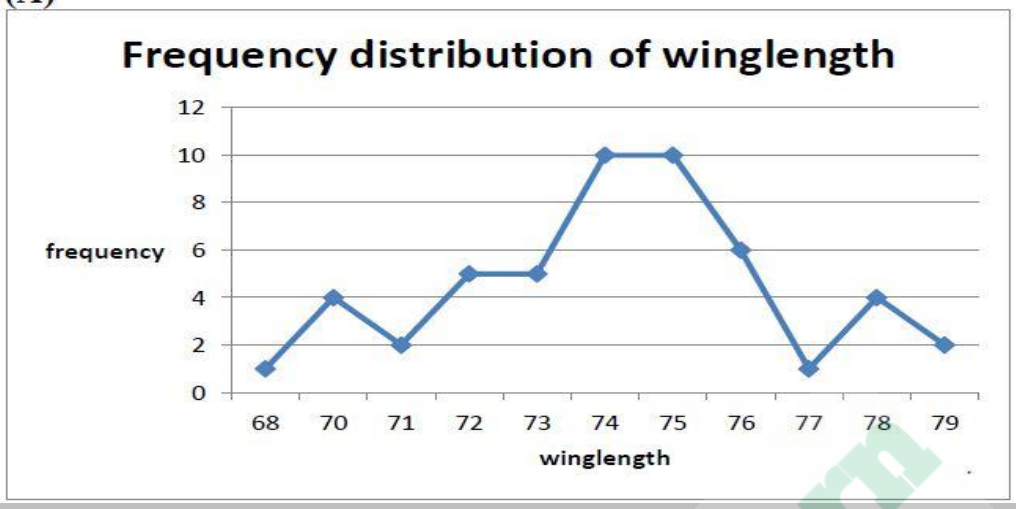
$\frac{f}{n} \times 100\%$
Where f = Frequency of the category class and
n = total number of values

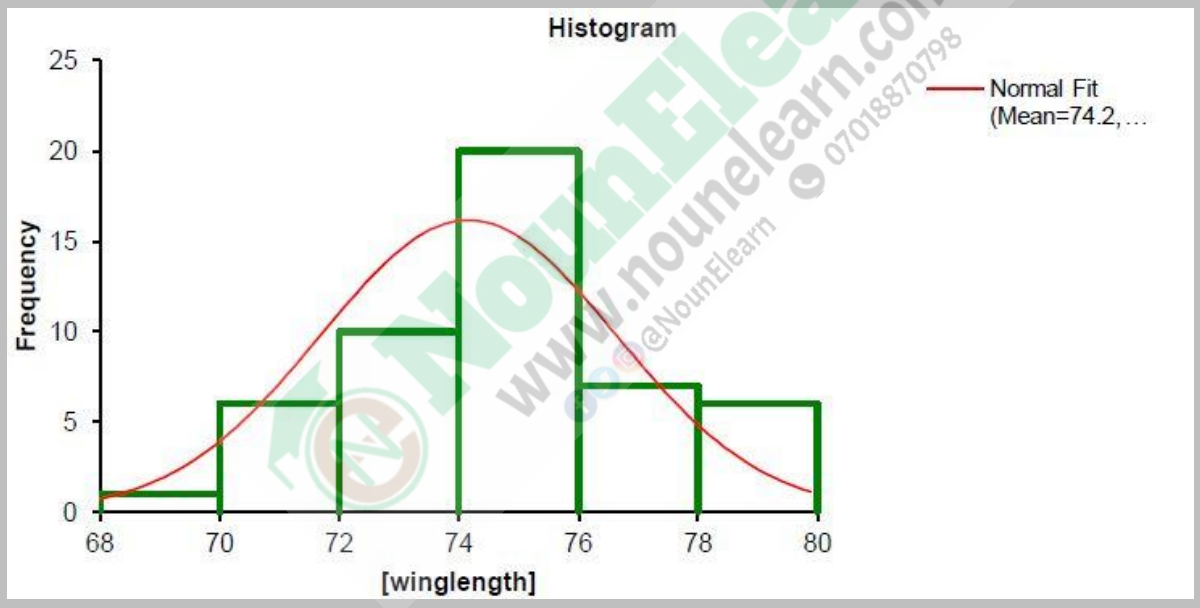
_____is a list of the figures in array form, occurring in the raw data, together with the frequency of each figure
Ungrouped frequency distribution

The range of set of data: *Highest value – lowest value* (79 – 68 =11)

Figures 2.1 (A-C): The frequency distribution of the wing length of 50 laughing doves

(A)





NB: Mean = 74.2

Calculations made from a grouped frequency distribution can never be exact (TRUE)

2.34 GROUPED FREQUENCY DISTRIBUTION

The heights in inches of commonly grown herbs are shown below. Organize the data into a frequency distribution with six classes, and make useful suggestions.

18	20	18	18	24	10	15	12	29	36
13	20	18	24	18	16	16	20	7	

Solution:

Find the range of the data: *Highest value – lowest value* (36 – 7 =29)

The class width is given by $\frac{\text{Range}}{\text{Number of classes}} = \frac{29}{6} = 4.8$ (round it up to the nearest whole number= 5)

The reasons for constructing a frequency distribution are:

1. To organize the data in a meaningful, intelligible way
2. To enable the reader to determine the nature and shape of the distribution
3. To facilitate computational procedures for measures of average and spread
4. To enable the researcher to draw charts and graphs for the presentation of data
5. To enable the reader to make comparisons among different data sets

Figure 2.3: A bar chart of the nutrient content of a seed

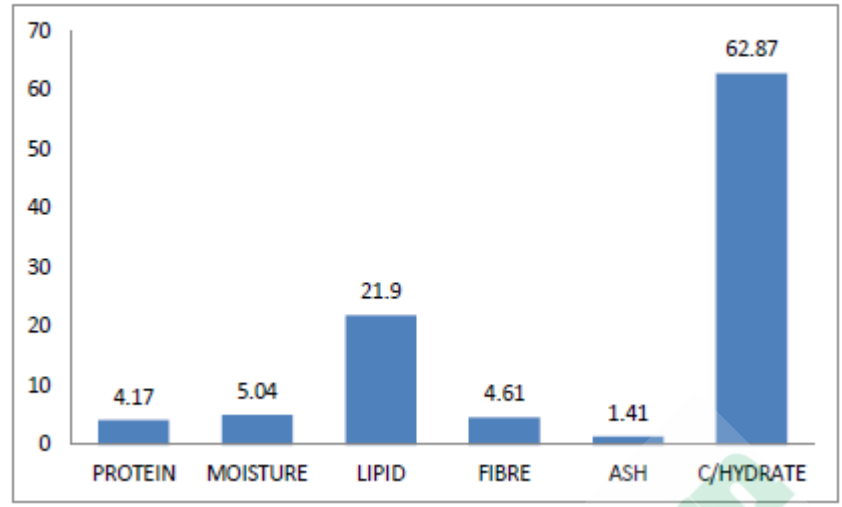
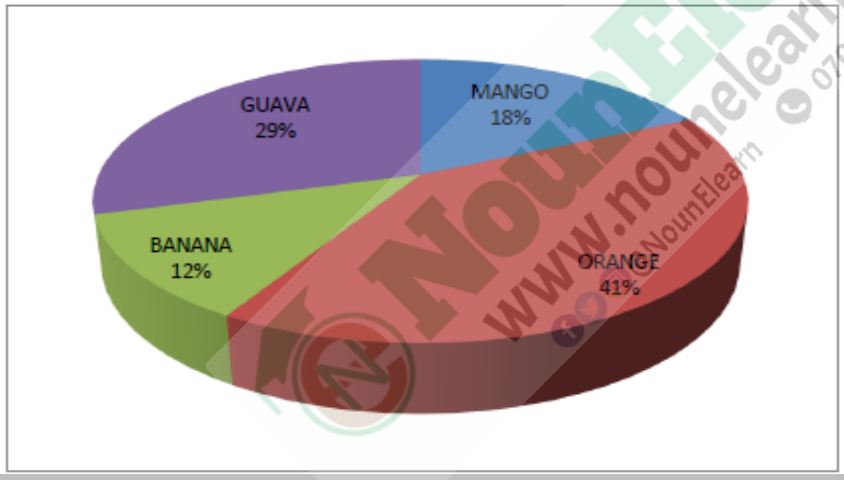


Figure 2.6: A Pie chart showing the distribution of fruits in an orchard



Forms of data representations include the following:

- bar chart
- cummulative frequency
- scatter plot or dot diagram
- pie chart
- the frequency distribution
- histogram

The types of frequency distributions are these

- ungrouped frequency distribution
- categorical frequency distribution

_____ is a branch of mathematics which can be defined as the chance of an event occurring
Probability

_____ is the basis of inferential statistics
Probability

Probability is the basis of _____
inferential statistics

A _____ is a scatter of related values
distribution

A _____is a scatter of related values, (such as the assortment of weights in a group of cattle)
distribution

The sum of a probability distribution is equals to_____
1

_____ is a complete listing of all possible outcomes of an experiment, together with their probabilities
P**robability distribution**

The most important distribution in statistics is called_____
normal distribution

Normal distribution is also known as _____
Probability distribution

The normal distribution is defined by two statistics _____
mean and standard deviation

The curve given by probabilities distribution approximates very closely to a Mathematical curve called_____
normal curve

The following are properties of a Normal Curve

- It is a unimodal symmetrical curve
- The mean, mode & median all coincide, thereby dividing the curve into two equal parts
- Most items on the curve are clustered around the mean
- No kurtosis or skewness in the curve

- The area beneath the curve is proportional to the observation associate with the part

Any value of an observation *X* on the baseline of a normal curve standardized as a number of standard deviation units, while the observation is away from the population mean, μ is called a *z*-score

To transform *x* into *z* the formula is given by:

$$z = \frac{(x - \mu)}{\sigma}$$

If the population mean μ of a normal curve is larger than the sample mean *x*, the *z* is **negative**

A Poisson distribution is a discrete probability distribution that is useful when **n** is larger and **p** is small

Poisson distribution is used to describe the spread of organisms like insects, trees, and snails' etc. by the following:
1. Divide the large area into small squares of equal size
2. Count the particular animal or plant species under study in each square
3. You can also randomly select a number of squares, if the area is two large

A distribution used to describe the spread of organisms like insects, trees, and snails is known as_____
Poisson distribution

A _____is a special probability distribution that describes the distribution of probabilities when there are only two possible outcomes for each trial of an experiment
binomial distribution

A binomial distribution is_____
a special probability distribution that describes the distribution of probabilities when there are only two possible outcomes for each trial of an experiment

_____ is the entire process of using an estimator to produce an estimate of a parameter
Estimation

Estimation and hypothesis testing are interrelated
True

An estimate is any specific value of a statistic while an estimator is any statistic used to estimate a parameter
True

A _____ is obtained when a single number is used to estimate a population parameter
point estimate

A _____ is a conjecture about a population parameter which may be true or false
statistical hypothesis

There are _____ types of statistical hypotheses
two

The two types of statistical hypotheses are _____ and **alternative hypothesis**
null hypothesis

The two types of statistical hypotheses are null hypothesis and _____
alternative hypothesis

The two types of statistical hypotheses are_____
null hypothesis and **alternative hypothesis**

The null hypothesis is symbolized by _____
Ho

A statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters is called_____
null hypothesis

Null hypothesis can be accepted or rejected
True

If the sample results do not support the null hypothesis then the conclusion which is on the rejection of the null hypothesis is known as the alternative hypothesis

Alternative Hypothesis is symbolized by_____
H1

The _____ states that there is a difference between a parameter and a specific value under study

null hypothesis

A null hypothesis is rejected when_____
the test value is in the critical region

The level of probability associated with an interval estimate is known as the _____
confidence level

The _____ is the degree of difference between sample mean (x) and population mean (μ*Ho*) that leads to the rejection of the null hypothesis
significance difference

The _____ is the range of values of the test values that indicates that there is significant difference and that the null hypothesis should be rejected
critical or rejection region

There are _____ possible outcomes in hypothesis testing
four

A _____occurs if one rejects the null hypothesis when it is true or it is the risk that a true hypothesis will be rejected
Type I error

A Type I error occurs if one rejects the null hypothesis when_____
it is true or it is the risk that a true hypothesis will be rejected

A Type II error occurs if you do not reject the null hypothesis when_____
it is false or when a false hypothesis is erroneously accepted as true

A _____ occurs if you do not reject the null hypothesis when it is false or when a false hypothesis is erroneously accepted as true
Type II error

The *t*-test is used when _____
the population standard deviation is unknown and the sample size is less than 30

The *t*-test is defined as a statistical test for the mean of a population and is used when the population is normally or approximately normally distribution, σ is unknown, and $n < 30$

Formula for *t*-test: $t_{n-1} = \frac{x - \mu}{s/\sqrt{n}}$

Where: x = Sample mean.
μ = Population mean.
s = Sample standard deviation.
n = Sample size.

Testing of hypotheses using the t-test follows the same procedure as for the z-test, except that you use the t- table instead.

_____ is the general method for testing compatibility based on a measure of the extent to which the observed and expected frequencies agree
Chi-square

Chi-square is also referred to as test for homogeneity randomness, association, independence and goodness of fit

In_____, k denotes the number of cells
chi-square distribution

In chi-square distribution, _____ denotes the number of cells
k

In chi-square distribution, k denotes the number of _____
cells

_____ is used to describe the spread of organisms like insects, trees, and snails' etc
Poisson distribution

A _____ is a discrete probability distribution that is useful when n is larger and p is small and when the independent variables occur over a period of time.
Poisson distribution

A binomial experiment is a probability experiment that satisfies the following four requirements:
1. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. i.e these outcomes can either be success or failure. No two events can occur simultaneously.
2. There must be a fixed number of trails.
3. The outcomes of each trial must be independent of each other.
4. The probability of a success must remain the same for each trial.

A _____ is a special probability distribution that describes the distribution of probabilities when there are only two possible outcomes for each trial of an experiment.
binomial distribution

_____ are interested in answering many types of questions.
Researchers

A _____ might want to know whether a new medication will lower a person's blood pressure.
physician

An _____ might wish to see whether a new teaching technique is better than a traditional one. **educator**

_____ is the entire process of using an estimator to produce an estimate of a parameter.
Estimation

_____ and hypothesis testing are interrelated.
Estimation

Estimation and _____testing are interrelated.
hypothesis

A _____ is obtained when a single number is used to estimate a population parameter.
Point estimate

A _____ is a conjecture about a population parameter.
statistical hypothesis

When you reject the Null Hypothesis, the five possibilities are:
1. There is direct cause and effect between the variables.
2. There is a reverse cause and effect relationship between the variables.
3. The relationship between the variables may be caused by a third variable
4. There may be a complexity of interrelationships among many variables.
5. The relationship may be coincidental.

In hypothesis testing situation, there are _____ possible outcomes.
four

_____ are usually constructed for the purpose of studying the relationship between two or more variables of classification.
Contingency tables

_____ tables are usually constructed for the purpose of studying the relationship between two or more variables of classification.
Contingency

_____ are used in a procedure that involves the comparison of the differences between the sample frequencies of occurrences or percentages that are actually observed and the hypothetical or theoretical population frequencies of occurrences or percentages that are expected if the hypothesis is true.
Chi-square distributions

Steps in the general x 2 testing procedure
1. Formulate the null and alternative hypotheses.
2. Select the level of significance to be used in the particular testing situation.
3. Take random samples from the populations, and record the observed frequencies that are actually obtained.
4. Compute the frequencies of percentages that would be expected if the null hypothesis is true.
5. Use the observed and the expected frequencies to compute the x 2 .
6. Compare the value of x 2 computed in step 5 with the x 2 table value at the specified level of significance (step 2).

_____and regression are other areas of inferential statistics which involve determining whether a relationship between two or more numerical or quantitative variables exists.
Correlation

Correlation and _____ are other areas of inferential statistics which involve determining whether a relationship between two or more numerical or quantitative variables exists.
regression

_____ is a statistical technique that measures the degree/strength of linear relationship in a bivariate normal distribution where two continuous variables drawn from the same population are compared.
Correlation

The _____ computed from a sample data measures the strength and direction of a linear relationship between two variables.
correlation coefficient

_____ is a statistical method used to describe the nature of the relationships between variables, that is, positive or negative linear or non linear.
Regression

A _____is a line that best fits the point in a scatter diagram, and it always passes through the point (X, Y).
regression line

Regression line is also called_____
least-square line

A _____ is the variable in regression that cannot be controlled or manipulated.
dependent variable (Y)

An _____ is the variable in regression that can be controlled or manipulated.
dependent variable (Y)

A _____ is a graph of the ordered pairs (x,y) of numbers consisting of the independent variable X and the dependent variable Y.
scatter diagram

_____ is a visual way to describe the nature of the relationship between x and y.
Scatter diagram

Regression can be _____or curvilinear
linear

Regression can be linear or_____
curvilinear

Regression can be_____
linear or curvilinear

_____ is when an increase in one variable is associated to a greater or lesser extent with an increase in the other.
Positive correlation

_____ is when an increase in one variable is associated to a greater or lesser extent with a decrease in the other.
Negative correlation

_____ is when a change in one variable is exactly matched by a change in the other variable.
Perfect correlation

_____ is when a change in one variable is almost exactly matched by a change in the other.
High correlation

_____ when a change in one variable is to a small extent matched by a change in the other.
Low correlation

_____ is when the two variables are not in matched at all, and there is no relationship between changes in one variable and changes in the other.
Zero correlation

_____ can be measured by finding the average of the products of the deviations of each of the paired variables from the overall mean of the relevant variable
Covariance

_____ deals with the methods of constructing and analyzing comparative experiments such as comparing the effect of different factors or treatments.
Experimental design

_____ ensures that maximum precision is achieved for the amount of effort expended in an experiment.
Experimental design

_____ is a systematic procedure for making observations under controlled conditions in such a way that they can be used for arriving at general conclusions regarding the population under study.
An experiment

An _____is a plant or animal or group of plants or animals making up a single replicate of a single treatment.
experimental unit

An experimental unit is also called a_____.
plot

_____ involves the allotting of treatments to the available material at random.
Randomization

_____ is the need for an experiment to be planned in such a way that it can be replicated or reproduced.
Replication / Reproducibility

CRD stands for_____
Completely Randomized Design

_____is the simplest type of experimental design in which treatments are assigned at random to a set of plots.

Completely Randomized Design (CRD)

_____ ensures that treatments allotment is completely random thereby avoiding biasness and minimizing inherent differences in experimental units or treatment.
CRD

Advantages of CRD
a. The design is very flexible and can be used for any number of treatments.
b. The statistical analysis is comparatively easy and straightforward.
c. It is unaffected by missing observations for any treatment for some purely random accidental reason.

The disadvantage of CRD is that _____
The design is inherently less informative than other more sophisticated layouts

_____ is an experimental design in which the total area is divided into blocks and all of the treatments are arranged within each block in a random order.
Randomized Block Design (RBD)

RBD stands for_____
Randomized Block Design

Advantages and Disadvantages of RBD
1. With heterogeneous material the residual variance can be reduced by choosing blocks of plots such that the plots within the blocks are fairly similar. i.e the design reduces the effect of heterogeneous material.
2. There is no restriction on the number of blocks or treatments, but in each block there must be the same number of plots, one to each treatment.
3. If some yields are accidentally lost, the analysis is again without due complications, although special modifications are required.

_____ is an experimental design in which the number of rows, columns, and treatments are equal and each treatment occurs just once in each row and column.
Latin square

_____ is a square design in that it consists of the same number of plots (though these don't have to be square) in two dimensions, i.e. 4 × 4, 5 × 5, 6 × 6, etc.
Latin square

A _____ is one where the treatments allocated to the experimental plots are combinations of two or more factors (hence the term "factorial").

factorial experiment

_____ is a systematic procedure for obtaining two or more estimate of variance and comparing them
Analysis of variance (abbreviated as ANOVA)

_____ as a technique enables the comparison of means and variances in an experiment that involves more than two treatments.
Anova

_____ estimates the value of the true variance of the population (σ2) from which the sample is drawn.
Anova

_____ permits us to conclude whether or not all means of the population under study are equal base upon the degree of variability in the sample data.
Anova

There are five advantages that non parametric methods have over parametric methods.
1. They can be used to test population when the variable is not normally distributed.
2. They can be used when the data are nominal or ordinal.
3. Can be used to test hypothesis that do not involve population parameters.
4. In most cases, computation is easier than in parametric.
5. They are easier to understand.

The following are the disadvantages of non-parametric test
1. They are less sensitive than in parametric i.e larger differences are needed before the null hypothesis can be rejected.
2. They tend to use less information than the parametric tests.
3. They are less efficient than their parametric counterparts when the assumptions of the parametric are met.

The_____, denoted by rs , is the nonparametric equivalent of the Pearson coefficient used for testing hypothesis when samples obtained are not normally distributed.
spearman rank correlation coefficient

_____is a statistical package developed especially for the social sciences but it has wide applicability in the areas of biology and agriculture.
SPSS

The _____ statistical software provides a wide range of statistical analysis and graphical capabilities.

MINITAB

_____ can be used for various types of statistical analysis ranging from editing and manipulating data, basic statistics, arithmetic, regression, Anova, non parametric test, exploratory data analysis et.c.
MINITAB

_____ estimate is obtained when a range of values is used to estimate a population parameter
Interval

The two statistical hypotheses for each situation are null hypothesis and the _____ hypothesis
alternative

_____is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters.
Null hypothesis

The _____is the degree of difference between sample mean (x) and population mean (μHo) that leads to the rejection of the null hypothesis.
significance difference

_____ error occurs if you do not reject the null hypothesis when it is false or when a false hypothesis is erroneously accepted as true.
Type II

When the population standard deviation is unknown and the sample size is less than 30, the type of test usually used is _____
t-test

By means of _____test it is possible to test the hypothesis that the two variables are independent
chi square

_____is the general method for testing compatibility based on a measure of the extent to which the observed and expected frequencies agree
Chi-square

_____ are areas of inferential statistics which determine whether a relationship between two or more numerical or quantitative variables exists.
Correlation and regression

The symbol use to represents samples correlation coefficient is _____

r

A regression line sometimes called the _____is a line that best fits the point in a scatter diagram, and it always passes through the point (X, Y).
least-square line

In the general equation for a fitted regression line Y=a+bX, Y is equal to ____ on the vertical axis.
Dependent variable

A _____is a graph of the ordered pairs (x,y) of numbers consisting of the independent variable X and the dependent variable Y.
scatter diagram

_____ correlation is when a change in one variable is exactly matched by a change in the other variable
Perfect

The full meaning of SPSS is _____
Statistical Package for Social Sciences

The type of test that is used if the population in a particular hypothesis-testing situation is not normally distributed is _____
Non-parametric tests

A stratified sample is a sample obtained by dividing the population into subgroups, called _____
Strata

In measuring the height of some rabbit, the mean was 76, and the standard deviation was 6, calculate the standard height of rabbits having a height of 112
6

The width of the termite house is normally distributed with mean 3cm and standard deviation 0.14. What width value separates the widest 10% of all such house from the other 90%?
3.179

A conjecture about a population parameter which may or may not be true is called_____.
Hypothesis

The area in the normal distribution diagram that indicate that a null hypothesis should be rejected is called the _____ region.
critical rejection

The risk of rejecting a true hypothesis is known as _____ error
Type 1

When a change in one variable is to a small extent matched by a change in the other, this is called____ correlation
Low

If the probability of a defective syringe is 0.4. The mean for the distribution of defective syringes in a total of 500 will____
200

A committee of 5 doctors can be chosen from 9 doctors in ____ ways
126

We have a _____ correlation when an increase in one variable is associated to a greater or lesser extent with a decrease in the other
negative

_____is the numeric location of the center of the class
Midpoint

_____is a discrete probability distribution that is useful when n is larger and p is small and when the independent variables occur over a period of time.
Poisson distribution

_____ is a special probability distribution that describes the distribution of probabilities when there are only two possible outcomes for each trial of an experiment.
Binomial distribution

Distribution that has no specific beginning or ending value is called _____ distribution
open ended

A scatter of related values, such as the assortment of weights in a group of chicken is called_____
distribution

_____ is simply a complete listing of all possible outcomes of an experiment, together with their probabilities.
Probability distribution

The normal distribution is defined by just two statistics, name them
mean and standard deviation

_____is a particular pattern of variation of numbers around the mean
Normal distribution

The mean yield of yam following the fertilizer treatment from 10 plots was 176.1kg with standard deviation 3.88. What is the 95% confidence limit for the mean yield of the yam?
(173.32, 178.88)

If 3% of the yam tuber harvest from a portion of farmland get rotten within 5 weeks of harvest, find the probability that in a sample of 100 tubers less than or equal to 2 tubers will be defective
0.4232

Adult males have normally distributed heights with mean equal to 70 inches and standard deviation equal to 3 inches, what percentage are between 68 and 73 in?
58.89%

Which of the following is not true about a binomial distribution?
The outcome of a new trial can depend on the trial of the previous

The width of the termite house is normally distributed with mean 3cm and standard deviation 0.14. What width value separates the widest 10% of all such house from the other 90%?
3.179

If a normal distribution has u=30 and sd(a)= 5, what is the 6th percentile of the distribution?
22.225

For a certain breed of monkey in a zoo, the time to start eating when food is supply is normally distributed with mean 1.25 min and standard deviation of 0.46 min. what is the probability that the time to start eating is between 1 min and 1.75min
0.5675

Evaluate P(-0.38<=z<=1.25)
0.5424

Which of the following is not true about a normal distribution curve?
Most items on the curve are clustered around the variance

In an experiment, the class with is 9, how many classes can be formed if the range of the set of data is 270.

30

A biologist discover that the proportion of 5 microorganisms in pond are V=24, W=36, X=15,Y=27, Z=38, determine the relative frequency of organism X
10.71

Which of the following is true of the projection of two vectors X and Y?
The projection of X on Y equal the scalar product of X.y where y is a unit vector in the direction of Y

Statistics that deals with using a conclusion deduce from a smaller sample to make conclusion on the entire larger population is defined as
Inferential

The heights in inches of tomato plants in a garden are shown below. If the data is organized into a frequency distribution with six classes, what will be the class width? 18 20 18 18 24 10 15 12 29 36 13 20 18 24 18 16 16 20 7
5

Information in the form of numerical figure is termed_____
data

In the construction of a frequency, which of the following is not important?
Give gaps in the classes to give room for lack of values.

Which of these is not a reason for constructing a frequency distribution?
So that drawing of charts and graphs may not be necessary

Which of the following is not a source of variation in a given population?
Type of habitat

Which of these is not frequently used as confidence levels employed in interval estimation?
98 percent

The critical or rejection region is_____
The range of values of the test values that indicates that there is significant difference and that the null **hypothesis should be rejected**

In poultry farming, the daily demand for water is normally distributed with a mean of 4000 litres and a standard deviation of 400 litres, if z-score of demand for a given day is 1.3, how many litres of water is the demand for the day?
4520 litres

The seven pairs of values (x, y) below shows the number of absences, x, in a BIO206 tutorial at Abuja study centre and the final exam grade, y, for 7 students. Find the correlation coefficient between x and y (1, 95), (0 , 90), (2,90), (6 ,55), (4, 70), (3 ,80), (3,85)
r = -0.93

The seven pairs of values (x, y) below shows the number of absences, x, in a BIO206 tutorial at Abuja study centre and the final exam grade, y, for 7 students. Find the equation of regression between x and y (1, 95), (0 , 90), (2,90), (6 ,55), (4, 70), (3 ,80), (3,85)
y = 3.56x - 9.57

Weight measurement is conducted on new harvest of Watermelon, Thirty (n=30) randomly selected watermelon are carefully weigh and the weight recorded, the mean weight of the sample is 28.6g and the sample standard deviation is 2.2g. Estimate a 95% confidence interval for the mean weight in the whole watermelon harvest
(27.81, 29.39)

Find the probability that seven of 10 persons will recover from a tropical disease if we can assume independence and the probability is 0.80 that any one of them will recover from the disease
0.2

The average number of customers arriving on any one day at poultry farming is 12, what is the probability that on a given day fewer than 9 customers will arrive at the farm?
0.16

Ten percent of the tools produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random exactly 2 will be defective by using poisson approximation to binomial distribution
0.18

Before applying a statistics in a research, which of the following statement is incorrect in what a researcher need to know?
Un-clarification of what to be achieved

If the probability of a defective syringe is 2/3. find the standard deviation for the distribution of defective syringes in a total of 729
12.73

Which of the following is not an experimentation principle?
diversion

A basket contains 5 oranges, 4 Mangoes, 3 Apples and 2 guavas. If 4 fruits are drawn at random, determine the probability that 1 of each fruit is drawn
0.12

Find the probability that in five tosses of a fair die a 5 appears at no time
3125/7776

A Biologist discovered that the proportion of 5 micro organisms in pond are V=24, W=36, X=15,Y=27, Z=38, determine the relative frequency of organism X
10.71

When do we have a negative correlation?
when an increase in one variable is associated to a greater or lesser extent with a decrease in the other

In a poultry farm, the daily demand for water is normally distributed with a mean of 4000 litres and a standard deviation of 400 litres, if z-score of demand for a given day is 1.3. How many litres of water is the demand for the day?
4520 litres

Information in the form of numerical figure is termed _____
Data

The statistical method in which quantitative information are gathered and handled is called _____
Descriptive

The type of variables that can be placed into distinct categories, according to some attribute is called _____
Qualitative

The subgroup of the population selected for study is referred to as _____
Sample

An estimation of the plant cover on a 5000 m by 5000 m land is to be conducted, this involves division of such land into smaller units and a number of such units are randomly selected and used as a sample. What type of sampling procedure is this? _____
Cluster

In a sampling procedure involving given fresh water habitat, 10 catfish was selected from the population of 100 catfish while 200 Tilapia fish was selected from a population of 2000 Tilapia fish. What type is sampling method is this?
Proportionate

What type of method is involved in a sampling procedure in which every fish in a fish pond has equal chances of being included in a sample?
Random

What type of method is involved in a sampling procedure in which every corn plant that is 1m apart in a particular farm will be included in a sample?
Systematic

The organisation of raw data in table form, using classes and frequencies is called _____
Frequency distribution

Find the mean of the following quantitative data:20 20 23 23 24 24 24 25 25 2527 27 27 27 2728 28 28 29 29 30 30 30 30 30 30 36 36 38 40
28

Find the mode of the following quantitative data:20 20 23 23 24 24 24 25 25 2527 27 27 27 2728 28 28 29 29 30 30 30 30 30 30 36 36 38 40
30

Find the class width for the data below:18 20 18 18 24 10 15 12 29 36 13 20 18 24 18 16 16 20 8
4.8

In a Mendellian crossing, what is the probability of getting a heterozygous tall plant in F1 plant?
0.5

What is the probability of obtaining a homozygous recessive plant in a Mendellian F1 plant?
0.00

x2 can take any value from zero to in infinity.
True

In a cross between a homozygous black sheep and homozygous white sheep, assuming the black colour is recessive, what is the probability of getting a white sheep in the F1?
1

In a cross between a homozygous black sheep and homozygous black sheep, assuming the black colour is recessive, what is the probability of getting a white sheep in the F2?
0.75

In a cross between a homozygous black sheep and homozygous black sheep, assuming the black colour is recessive, what is the probability of getting a black sheep in the F2?
0.25

Given the formula z = X-μs , what does μ stand for?
Population mean

Given the formula z = X-μs , what does X stand for?
Sample mean

A statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters is called _____
Null

The range of values of the test values that indicates that there is significant difference and that the null hypothesis should be rejected is termed _____
Critical values

What is the meaning of x2?
Chi-square

A general method for testing compatibility based on a measure of the extent to which the observed and expected frequencies agree is termed _____
Chi-square

Given the equation to test for relationship between variables as Y = a + bX, what does 'a' stands for?
Intercept

In a complete randomized design, what type of analysis is best suited for comparing the means?
One-way ANOVA

The non- parametric equivalent of the Pearson coefficient used for testing hypothesis when samples obtained are not normally distributed is called _____
Spearman rank correlation

When the data that are being analysed using statistical tools, are derived from the fields of biological sciences it is called _____
Biostatistics

A subgroup of the population selected for study is termed _____

Sample

A _____ sample is one that has the same chance as any other of being selected
Random

Calculate the percentage of B blood group among the 40 students in a Biostatistics class_____A, AB, B, O, O, A, B, AB, A, B, O, O, O, A, AB, B, B, A, O, AB, A, O, O, A, AB, B, B, A, A, B, AB, A, O, B, AB, O, A, B, A, B.
27.5

The full meaning of RBD is _____
Randomized Block Design

In a Spearman rank correlation, if the two sets of data have the same ranks, then rs will be _____
+1

Systematic procedure for making observations under controlled conditions in such a way that they can be used for arriving at general conclusions regarding the population under study is called (a)an _____
Experiment

The range of values for correlation coefficient is from _____
-1 to +1

Which of the following is not a common statistical package?
MINITAB

Which of the following analyses can be executed using SPSS?
All the given options

The full meaning of SPSS is_____
Statistical Package for Scientific Study

In a Randomized block design, what type of analysis is best suited for comparing the means?
Latin square

An experimental design in which the total area is divided into blocks and all of the treatments are arranged within each block in a random order is called_____
Completely randomized design

Which experimental design is best suited for evaluating the effects of a feed of different protein levels on the growth of quails or fish of the same age?

Completely randomized design

Find the correlation between the yield of pepper plant and the length of the plant in the following:
0.954

The principle involved in any experimentation include the following except_____
Randomization

What type of correlation is provided in the chart below?
Perfect positive correlation

In a situation whereby a change in variable A is almost exactly matched by a change in variable B implies_____
Zero correlation

Which of the following is not an example of a curvilinear regression?
Algorithmic

What type of correlation is involved in studying the relationship between growth in snails and different feed protein levels, quantity of feed given per day and hours of lighting per day?
Multiple relationship

The equation Y = a + bX represents_____
Correlation

A young researcher is interested in finding out the relationship between height and yield in okra plant, which of the following statistical analysis is best suited for such analysis?
Correlation

Which of the following is not true about Chi-square distribution?
It is symmetrical

Which of the following in true about the student's t-test?
All the given options

A Type II error occurs if_____
One does not reject the null hypothesis when it is untrue

Assuming the following values of a sample are given: x = 454. n = 120, standard deviation (S) = 27, μ = 460, α = 0.05 or 95 (confidence coeff.). What is the status of H0?
Accept

Which of the following statements is not true about statistics?
It refers to collection of quantitative information

Biostatistics is used in the following field of endeavours except
Biology

Which of the following is not a source of variation in a given population?
Age of the population

Which of the following is true about continuous variables?
Can only be assigned integers

The following are the reasons why sampling is important except
It saves time

Which of the following sampling methods is not part of the common types?
Random

The following are the advantages of using a cluster sample over other types of sampling methods except:
Cost reduction

Which of the following is the correct formula for calculating a class midpoint?
Xm= Lower Limit+Upper limit2

Which of the following is not a method of data representation?
Bar chart

In a study on the distribution of tree-roosting birds, if there are 200 birds randomly distributed on 500 trees, what is the probability that a given tree contains exactly three birds?
1

A binomial experiment usually satisfies the following requirements except
There must be a fixed number of trials

Which of the following symbols conforms to a null hypothesis?
H3

Which of the following is not true about a one-tailed test?
A one-tailed test right- tailed when the inequality sign is t

In a two-tailed test, when should the null hypothesis be rejected?
When the test value is in either of the two critical regions

Which experimental design is best suited for evaluating the effects of a feed of different protein levels on the growth of quails or fish of the same age?
Completely randomized design

In a complete randomized design, what type of analysis is best suited for comparing the means?
One-way ANOVA

In a situation whereby a change in variable A is almost exactly matched by a change in variable B implies_____
High correlation

Which of the following is not an example of a curvilinear regression?
Algorithmic

In a correlation analysis, it was observed that dose of gamma irradiation is negatively correlated with wing length of Drosophila melanogaster, interpret this observation.
The higher the dose, the shorter the wing

A young researcher is interested in finding out the relationship between height and yield in okra plant, which of the following statistical analysis is best suited for such analysis?
Correlation

A general method for testing compatibility based on a measure of the extent to which the observed and expected frequencies agree is termed _____
Chi-square

Which of the following is not true about Chi-square distribution?
It is symmetrical

Find the mode of the following quantitative data: 20 20 23 23 24 24 24 25 25 25 27 27 27 27 27 28 28 28 29 29 30 30 30 30 30 30 36 36 38 40. The answer is_____
30

Find the mean of the following quantitative data: 20 20 23 23 24 24 24 25 25 25 27 27 27 27 27 28 28 28 29 29 30 30 30 30 30 30 36 36 38 40. The answer is_____
28

Given below are the heights (cm) of 20 maize plants at 60 days after planting: 23, 50, 38, 42, 63, 75, 12, 33, 26, 39, 35, 47, 43, 52, 56, 59, 64, 77, 15, 21, 51, 54, 72, 68, 36, 65, 52, 60, 27, 34, 47, 48, 55, 58, 59, 62, 51, 48, 50, 41, 57, 65, 54, 43, 56, 44, 30, 46, 67, 53 What is the range? The answer is_____
65

The organisation of raw data in table form, using classes and frequencies is called_____
Frequency distribution

What type of method is involved in a sampling procedure in which every fish in a fish pond has equal chances of being included in a sample?
Systematic

In a sampling procedure involving given fresh water habitat, 10 catfish was selected from the population of 100 catfish while 200 Tilapia fish was selected from a population of 2000 Tilapia fish. What type is sampling method is this?
Proportionate sampling

The following are the advantages of using a cluster sample over other types of sampling methods except:
It is more accurate than other sampling methods

An estimation of the plant cover on a 5000 m by 5000 m land is to be conducted, this involves division of such land into smaller units and a number of such units are randomly selected and used as a sample. What type of sampling procedure is this?
Cluster

Which of the following sampling methods is not part of the common types?
Striated

The following are the reasons why samplings are important except
It gives the most accurate and exact information about a population under study

The subgroup of the population selected for study is referred to as_____
Sample

Which of the following is true about continuous variables?
They are obtained by measuring

A data obtained from measuring the weight of individuals in a population is an example of_____
Quantitative variables

A variable population classified in terms of their ability to roll the tongues is an example of_____
Qualitative variables

The type of variables that can be placed into distinct categories, according to some attribute is called_____
Qualitative variables

Which of the following is not a source of variation in a given population?
Age of the population

Biostatistics is used in the following field of endeavours except_____
Agriculture

The statistical method in which quantitative information are gathered and handled is called_____
Descriptive statistics

Which of the following statements is not true about statistics?
It does not need any critical observation before arriving at conclusion

Information in the form of numerical figure is termed_____
Data

NounElearn www.nounelearn.com @NounElearn