J BLACK EDU CONSULT

MOTTO: KNOWLEDGE FOR CAREER CONSCIOUSNESS

FOR MORE SUMMARY CONTACT ME ON:

WHATSAPP: 08122993587/08063755610

STT102: INTRODUCTORY STATISTICS

_____is the science that deals with the collection and summarization of facts that are expressible in numerical forms

Statistics

_____ involves the comparison and summarization of facts

Statistic

Statistical questions include questions like:

How many Nigerian mothers wine their babies at nine months' old?

What is the life expectancy of a Nigerian woman?

_____ is defined as the part of decision making which relates to numerical information

Statistics

_____ is the collection of all individuals, items, or data under consideration in a statistical study

Population

J BLACK EDUCATIONAL CONSULT

There are _____ types of population

Four (4)

The types of population include:

- **Finite population**
- **Infinite population**
- **Countable population and**
- **Uncountable population**

_____ is obtained by tossing a coin, throwing a die or drawing slips from a box

A random sample

Characteristics possessed by members of a population is called_____

A variable

A variable could take _____or _____ values

Integral or real

Specific values are called_____

Discrete variables

Unspecific values are called_____

Continuous variable

_____investigates the existence of relationship between two variables or characteristics

Statistical study

The determination of the relationship between smoking habit and lung cancer or between vasectomies and prostate cancer both requires_____-

Statistical methodology

Statistical methodology has _____ branches

Two (2)

The two statistical methods include:

- **Descriptive statistics and**
- **Inferential statistics**

_____ is concerned with the organization, presentation, and summarization of data

Descriptive statistics

J BLACK EDUCATIONAL CONSULT

_____ involves procedures employed to draw conclusion about a large body of data called population

Inferential statistics

The population in a collections of data is usually based on a smaller set of data called_____

Sample

_____ is a more common place method of gathering information

Sampling

There are _____ types of random samples or sampling

Five (5)

The two types of random samples or sampling includes:

- **Simple random sample**
- **Systematic random sample**
- **Cluster random sample**
- **Stratified random sample**
- **Multistage random sample**

_____ is a sample obtained by simple random sampling

Simple random sample

A process for which each possible sample is equally likely to be selected is called _____

A simple random sampling process

There are _____ types of simple random sampling

Two (2)

- **Simple random sampling with replacement**
- **Simple random sampling without replacement**

_____ is a sample where a member of the population is selected more than once

Simple random sampling with replacement

_____ is a sample where a member of the population is selected only once

Simple random sampling without replacement

_____is one of the methods of obtaining a simple random sample

A table of random numbers

_____is a computer package used to obtain a simple random sample from a population

J BLACK EDUCATIONAL CONSULT

Minitab

_____is the command used in a computer system to perform a simple random process

Sample

SOME SHORTCOMINGS OF A SIMPLE RANDOM SAMPLE INCLUDE;

1.**when the population being sampled is larger it is not practical to obtain a simple random sample by picking slips of paper out of a container**

2.failure to provide sufficient coverage when information about sub-population is required

3.its impracticality when the members of the population are not contiguous

_____sampling is easier to implement than simple random sampling

Systematic random sample

The process of systematic random sample includes:

1. **The population size is divided by the sample aize and the result is rounded up to the nearest whole number say k**
2. **Obtain a number a, say, from a random number table (or from a simple device**
3. **A must lie between I and n**
4. **Select for the sample a, a+k, a+2k, which are members of the population.**

_____ method is useful in the case when members of the population under consideration are widely scattered geographically

Cluster sample

Cluster sampling can be executed in these following steps:

1. **The population is divided into groups (clusters)**
2. **A simple random sample of the clusters is obtained**
3. **The sample is then taken as all of the members of the clusters obtained in.**

Clusters does not mirror the entire population and members of a cluster are frequently more homogeneous than the members of the population as a whole is an advantage of _____

Cluster sampling

_____ is obtained by dividing the population into sub-populations (or strata) and then sampling is done from each stratum.

Stratified sampling

Each stratum is homogeneously related to the characteristics under consideration is an advantage of _____

Stratified sampling

J BLACK EDUCATIONAL CONSULT

Stratified sampling method is often more reliable that _____

Cluster sampling method

_____ is the combination of all the earlier mentioned sampling processes

Multistage sampling

_____ sampling method is usually used by government agencies

Multistage sampling

_____ sampling method can be used by federal ministry of health on the population to acquire information on illnesses, injuries, mortality rate and other health issues

Multistage sample

_____ is the purpose of sampling

To obtain study classes that are representative of the population of interest

The use of _____ eliminates investigators-induced bias in the sample selection process

Random samples

_____ sampling schemes are costly, time-consuming, difficult or impossible to carryout

Random

A sample from a population of observation is said to be _____ if the sample is not representative of the parent population from which it is part

Bias

There are _____ ways of representing data

Four (4)

- **Tabulation**
- **Diagrams, charts, and graphs**
- **Pictograms**
- **Block and bar diagrams**

To simplify and facilitate the interpretation between trends and features arising from the statistical investigation_____ is needed

Tabulating (tabulation)

J BLACK EDUCATIONAL CONSULT

Pass Percentage of BA and B.Sc. (2014-15)			
	No. of Students Appeared in the Annual Examination	Passed	Pass Percentage (%)
Botany	29	24	82.76
Chemistry	30	25	83.33
History	40	34	85.00
Malayalam	38	37	97.37
Mathematics	32	28	87.50
Physics	28	17	60.71
Zoology	30	23	76.67

A TABLE

There are _____ types of tables

Three (3)

The three types of tables include:

1.source or reference tables

2.working table

3.summary table

_____and_____ description are carried out after tabulating a collection of data

Pictorial and diagrammatic description

_____is the simplest method of presentation of information

Pictorial figures

_____ are commonly used in advertisements and in printing media

Pictorial figures

_____ are used to convey broad relationships involving two or more variables

Pictorial figures

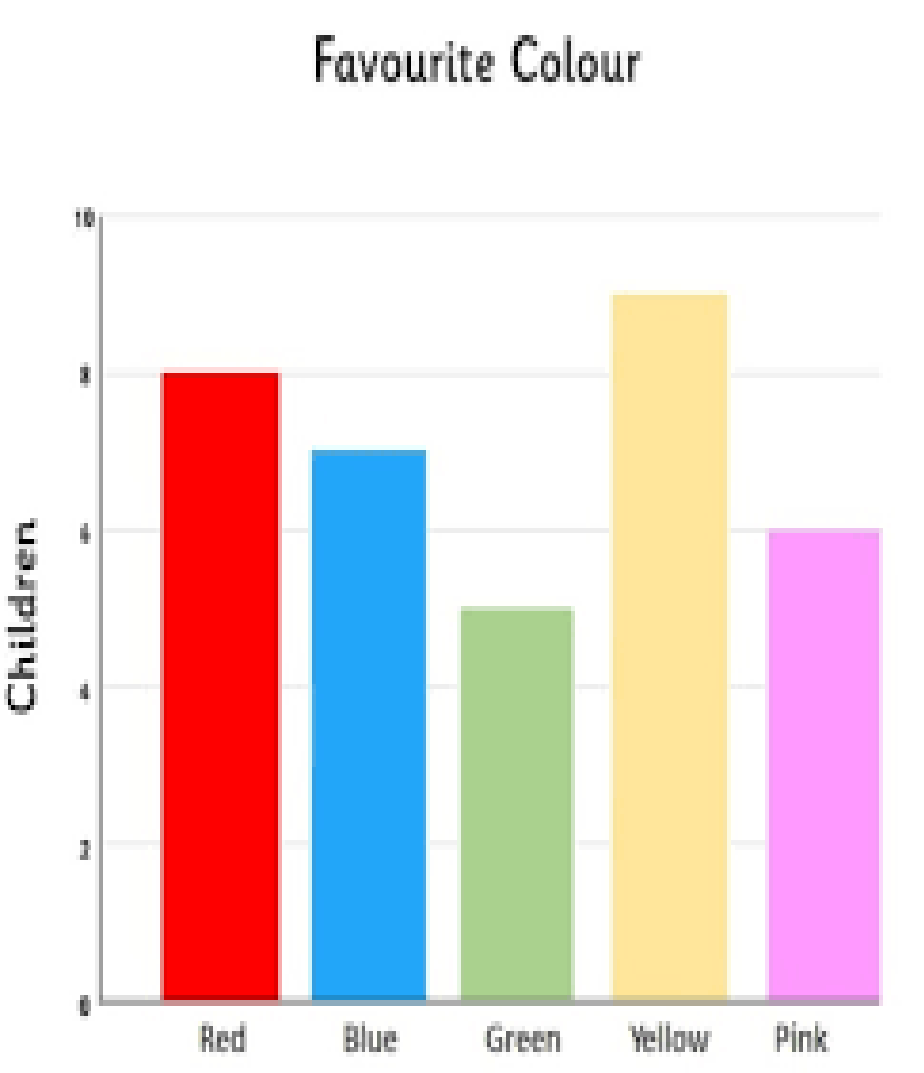
_____ consist of a line, bar or column of equal thickness but variable lengths

Bar chart (diagram)

A bar diagram or chart could be _____or _____

Vertical or horizontal

J BLACK EDUCATIONAL CONSULT



A VERTICAL BAR CHART

A chart divided into pie-shaped pieces proportional to its relative size is called _____

Pie chart

Raw data resulting from a sample survey or a census are usually in the form of _____ values

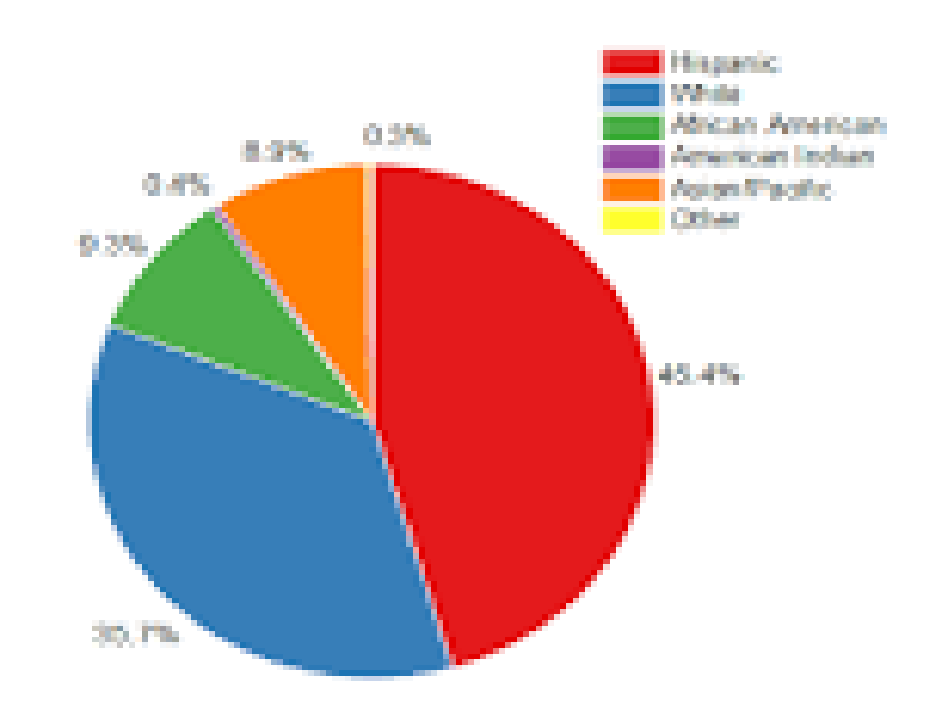
Unorganized

In other to make important features obvious and apparent a set of data must be ____and _____

Organized and summarized

The number times each score in a collection of data occurs in the sample is called_____

Frequency



A PIE CHART

The table used to display each frequency score is termed_____

Frequency distribution

The comparison of two different sets of data is usually done through the proportion or percentage of the total number of observation falling into each interval which is known as _____

Relative frequency OR

_____is the proportion or percent of the total number of observations falling into each interval in a frequency table

Relative frequency

$$\text{Relative frequency} = \frac{\text{Absolute frequency}}{\text{total number of observation}}$$

_____is a graphical representation of either frequency or relative frequency tables

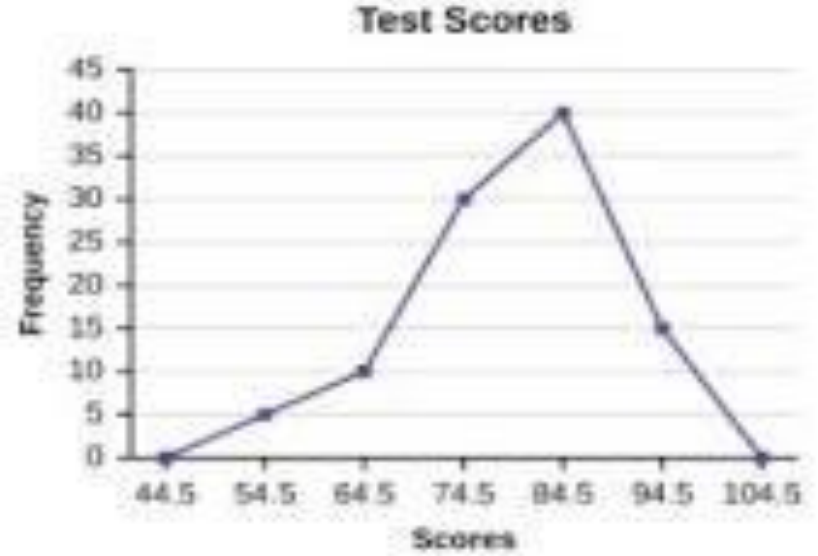
Histogram

To construct _____ you must place the true class intervals on the horizontal scale and the frequencies or the relative frequencies on the vertical scale ensuring that the height of each bar is equal to the frequency (or relative frequency) of each interval

Histogram

A_____ like a histogram is a graphical representation of a frequency table

Frequency polygon



A. Frequency polygon

_____distribution is one in which the frequencies are equal or almost equal to 80.95 on either side of a center point

A symmetric

_____distribution is one in which the frequencies are unequal on either side of a center point

Non symmetric,

The method of showing the classes on the horizontal axis of a histogram is called_____

Class boundaries

_____is the number half way between the lower class limit of the class and the upper class limit of the next lower class

Lower class boundary

The number half way between the upper class limit of the class and the lower class limit of the next higher class is called_____

Upper class boundary

For instance, consider the classes 40-49, 50-59 and 59-60 of the ages of 40 cancer patients given below (see Table 7.1). We have that

Ages of patients	*Frequency (No. of patients)*	*Relative frequency*
30-39	3	0.075
40-49	1	0.025
50-59	8	0.200
60-69	10	0.250
70-79	7	0.175
80-89	7	0.175
90-99	4	0.100
	40	1.000

Lower class boundary $= \frac{49+50}{2} = 49.5$

Upper class boundary $= \frac{59+60}{2} = 59.5$

_____is obtained by running the frequency of all classes representing values less than the specified class limit.

Cumulative frequency

_____is obtained by dividing the corresponding cumulative frequency by the total number of pieces of data

Cumulative relative frequency

A graph of the cumulative frequency distribution for any collection of data is called_____

Cumulative frequency curve

Cumulative frequency is also called _____

Ogive

· Cumulative relative frequency is $\frac{4}{40} = 0.10$

You may interpret this information by saying that 10% of the cancer patients are less than 50 years.

The middlemost value about which other values cluster is called _____

Measure of location or central tendency

_____is the most commonly used measure of tendency

Mean

_____is the arithmetic average of the collection of data

Mean OR

The sum of the collection of data is called _____

Mean

Mean is denoted with the symbol

$$\overline{X} = \frac{\sum X}{N}$$

EXAMPLE: find the mean of the following; 25, 27, 28, 30, 35,

$$\overline{X} = \frac{X_1 + X_2 + X_3 \ldots X_N}{N}$$

Where
$\overline{X}$ = the mean
X_1 = the first value
X_2 = the second value
X_3 = the third value
X_N = the last value
N = the number of valuse

Mean x= 25+27+28+30+35 =29

5

The mean is the most useful because it is more amenable to arithmetic manipulations true or false?

True

_____is the middle number in the list of numbers

Median

To find the median of any set of numbers the numbers must be arrange in either _____or _____ order

Ascending (from small to big) or descending (big to small)

Example: 12, 13, 14, 15, 16, 17, 18,

J BLACK EDUCATIONAL CONSULT

From here the median is _____

15

But if u have two middle number

E.g.; 12, 13, 14, 15, 16, 17, 18, 20.

To get the median here add the two middle numbers and divide by two

Mean= $\frac{15+16}{2}$ =15.5

_____ is preferable to use when the data has the possibility of extreme value

Midian

The _____ is the most frequently occurring number

Mode

Eg; 15, 15, 15, 12,13,14,16,12,13

The mode is 15 since this is the most frequently occurring score or value

_____is more useful when description of qualitative data is required

The mode

There are _____ measures of dispersion

Three (3)

The measures of dispersion include:

- **Range**
- **Variance**
- **Standard deviation**

_____is the difference between the smallest and the largest score in the sample

Range

E.g. 30,40,50,60,70,80,

Range = 80-30 =50

_____ is defined as the average of the squared difference from the mean

Variance

The distance between a measure in the set and the mean value for the set is called _____

Standard deviation

J BLACK EDUCATIONAL CONSULT

---------- shows the relative positon of any individual score with respect to all scores in the collection

Percentile

Percentile= number of scores less than given scores x 100

Total number of scores

If 80 nurses took the qualifying examination and 16 nurses scored lower than nurse H, percentile score of x is given by

16 x 100 = 20%

80

But if 64 nurses had scores lower than H, nurse H percentile would be

64 x 100 = 80%

80

This means that a percentile score of 80 show that nurse H is in the top 20% of the group

_____is a numerical means which facilitates the comparison of two scores or sets of scores

Percentile

How to differentiate the relationship between two or more variables is a problem faced in_____?

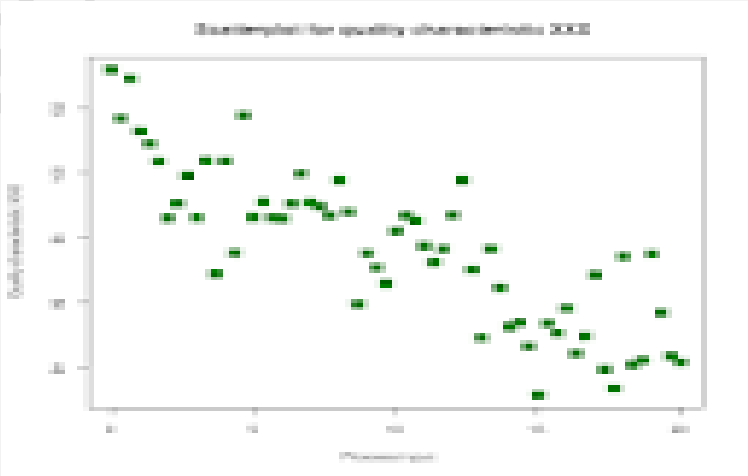
Statistics

In the correlation study data are usually arranged in _____

Pairs

-_____gives you the visual image of the relationship studied

A scatter diagram



A scatter diagram

J BLACK EDUCATIONAL CONSULT

The numerical representation of the relationships between variable is called_____

A scatter diagram involves plotting each pair of _____ as a single point

Observations (x1,x2)

Correlation coefficient (rs)

There are _____ types of correlation

Two (2)

_____ and _____ are the types of correlation coefficient

Product moment correlation coefficient and spearman's rank order correlation coefficient

_____is a non – parametric or distribution free statistical technique

Spearman's rank order correlation coefficient

When the assumptions underlying the classical techniques fails_____is used

Spearman's rank order correlation coefficient

It is more convenient to use _____ when data values are assigned ranks

Spearman's rank order correlation coefficient

----------is used when the relationship between two continuous variables is curvilinear

Correlation ration(h)

If the points appear to lie on a curve the relationship is_____

Curvilinear

In correlation study the initial step is_____

The display of a scatter diagram

_____are index numbers whose values range from -1 to +1

Correlation coefficient

Correlation does not mean_____

Causation

----------is determined by any two distinct points that lie on that line

A straight line

_____is the concept behind finding the regression line

The assumption that data points are scattered about a straight line

_____is defined as the likelihood of occurrence of an event

J BLACK EDUCATIONAL CONSULT

Probability

the number of ways of achieving success

Probability = the total number of possible outcomes

_____is defined as a specific result which has the likelihood of occurrence or not in an experiment

An event OR

_____ consist of a collection of outcomes

An event

Any probability very close to _____is an indication that the event is unlikely to occur

zero

a probability of _____indicates that the event will surly occur

1 or 100%

_____is a model where by possible outcomes of an experiment are equality likely to occur

Equal likelihood model

Probability of an event = F

N

NOTEABLE PROPERTIES OF PROBABILITY INCLUDES;

Probability of an event lies between _____and_____inclusive

0 and 1

Probability of an event not occurring is _____

Zero (impossible event)

Probability of an event occurring is _____

1 (certainly occurring)

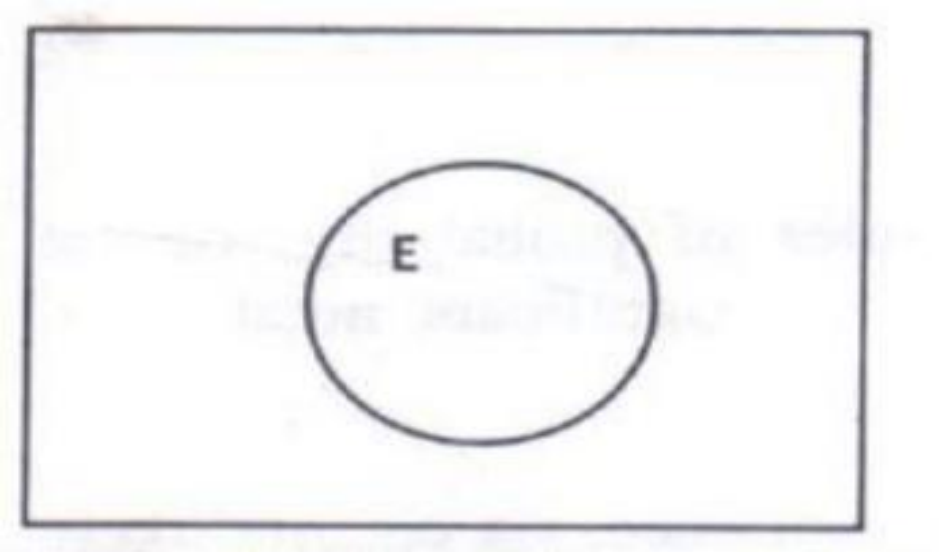
_____is the conventional method of representing events

The use of letters A, B, C, D

J BLACK EDUCATIONAL CONSULT

Visual representation of events and relationships between them is done by means of _____

Venn diagrams



Venn diagram for event

The sample space as shown above is shown as_____

A rectangle

RULES OF PROBABILTY

Special addition rule

Complementation rule

General addition rule

Multiplication rule

_____ is defined as that probability that the event occurs under the assumption that another event has occurred

Conditional probability

Two events are said to be _____if the occurrence or nonoccurrence of one of the events does not affect the probability of the other event

statistically independent

meaning that event B is said to be statistically independent of event A if the occurrence of event A does not affect the probability that event B occurs.

J BLACK EDUCATIONAL CONSULT

----------is a set of all measurements of interest

Population

_____is a Population to be investigated

Target population

_____ is a population whose conclusion are yet to be drawn

Sample population

_____is just a part of the population

Sample

_____ is referred to as a process by which you draw conclusion about a population given the information contained in a sample

Statistical inference

When conducting a research, the value of the population parameters is _____ since a part of the **population is observed**

Distributions are classified into _____ general patterns

Two (2)

Empirical and

Theoretical distribution

And empirical distribution is one obtained from_____
tabulating a frequency distribution from a set of sample measurements.

_____is the frequency for all the measurements in an entire population

Theoretical distribution

_____is rarely obtainable in practice

Theoretical distribution

A particular shape of continuous symmetric distribution is called _____

Normal distribution

Frequency distribution could be _____

Continuous or discontinuous and symmetrical or skew

A sample to acquire more information about a population is more preferable than_____

Taking census

An anticipated amount of error in a sampling procedure is called_____

J BLACK EDUCATIONAL CONSULT

Sampling error

The central limit theorem states that;

for any given population, not necessarily normal distributed and having mean and standard deviation, the sampling for fixed n which is generated from this population will be approximately normally distributed with mean and standard deviation.

There are _____aspects of statistical inference

Two (2)

_____ and_____ are types of statistical inference

Estimation and hypothesis testing

Results of a hypothesis test is said to be _____if the null hypothesis is rejected at a chosen level

Statistically significant

A method of estimating population parameters using information in the sample

Is called_____

The confidence interval procedure

_____is employed in making decision or passing judgments about the value of a parameter

Statistical inference

_____is one of the most widely used methods in decision making Process

Hypothesis test

_____is a statement that something is true

Hypothesis test

A hypothesis test involves _____

Two mutual exclusive hypothesis

_____and _____ are mutual exclusive hypothesis

Null and alternative hypothesis

_____is the first step in hypothesis test

Decide on the null and alternative hypothesis

Taking the Statistics of _____present her substantial problems in comparison to statistics of death

J BLACK EDUCATIONAL CONSULT

Sickness

This because death is a unique event, which occurs at a point in time whereas the occurrence of illness may be repeated in the same person from same or different causes, and it'd occurrence is for a period of time.

Death is defined in contrast to _____

Illness

J BLACK EDUCATIONAL CONSULT