BUSINESS STATISTICS STT 060

1.0 INTRODUCTION

1.1 Meaning of Statistics

- Statistics can be defined as the collection, presentation and interpretation of numerical data- Croxton and Crowed
- 2. Statistics are numerical statement of facts in any department of enquiry placed interrelation to each other- Bouly
- 3. Statistics are measurements, enumerations or estimates of natural or social phenomena in a systematic arrangement to exhibit their inner relation- Conner
- 4. By Statistics we mean quantitive data affected to a marked extent by a multiplicity of causes- Youle and Kendal
- 5. The Science of Statistics is essentially a branch of applied mathematics and can be regarded to as a mathematics applied to observation data. Rev Eister
- 1.2 <u>The Scope and Importance of States is</u>
 - 1. *Statistics and Planning* Statedes is indispensed into panning in the modern age which is termedeas the age of planning comost all over the world the governments are going back to planning for economic development
 - 2. *Statistics and economics*: Statistical data and techniques of statistical analysis are immensely involved in solving economical problems. Such as wages, price, time series analysis and demand analysis
 - 3. *Statistics and Business*: Statistics is major tool applied in production control. Statistical techniques are being used more in studying the desires of the customers
 - 4. Statistics and Industry: Statistics is used in inequality control. This is in Production engineering to find out whether the product is confirming to the specifications or not. Inspection plan and control chart are the major statistical tools applied.
 - 5. *Statistics and Mathematics:* The recent advancements in statistical techniques are the outcome of wide applications of mathematics

• Useful in gaining insight into topics difficult to gather info through the other methods

Disadvantages

- Susceptible to facilitator bias
- Discussions can be dominated by few individuals
- Data analysis is time consuming
- Doesn't provide valid information at the individual level
- The information is not representative of other groups •

5. Surveys and Questionnaires

Advantages:

- Administration is comparatively inexpensive •
- Reduces chances of evaluator bias
- Many people are familiar with surveys
- Many people comfortable with a survey rather than interview
- Tabulation of Close- ended responses is an easy and straight forward proces. Vantages: Respondents may not respond leading between the straight forward process. ٠

Disadvantages:

- Respondents may not respond leading •
- Items may not have the sam the nearing to all rea
- Size and diversiv or sample will be multiply by people's ability to read •
- Character of contact with reasonant, you never know really who completed the survey
- Unable to probe for additional details •
- Good survey questions are hard to write •

Other methods include

- 6. Town hall meetings and other large group events
- 7. Case studies
- 8. Illustrated Presentations- Photo Voice, Power Voice.

4.2 THE MEDIAN

The Median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

EXAMPLE:

Find the Median for the following data

25,18,27,10,8,30,42,20,53

Solution: Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

When odd number values are given, The middle value is the 5th item .i.e. 25 is the median

Using the Formula



When even numbers are given....Find the median for the following data

5,8,12,30,18,10,2,22

Arranging the data in the increasing order 2,5,8,10,12,18,22,30

Here the median is the middle two items i.e mean of (10, 12) i.e

$$= \left(\frac{10+12}{2}\right) = 11$$

= Using Formula = $\left(\frac{9}{2}\right)$ th
item= 4.5th item = 10 + (1/2) (12-10)
= 10 + $\left(\frac{1}{2}\right)$ x 2 = 11

 f_{1} = frequency of the modal class

- f_0 = frequency of the class preceding the modal class
- f_2 = frequency of the class succeeding the modal class

The above formula can also be rewritten as:

Mode= L +
$$\underline{f_1 - f_0}_{2 f_1 - f_0 - f_2} X C$$

4.4THE RELATIONSHIP BETWEEN THE MEAN, MEDIAN AND THE MODE:

If in a distribution mean= median= mode, then that distribution is known as symmetrical distribution. If mean \neq median \neq mode, then it is a skewed distribution and such a distribution could either be positively or negatively skewed.

(a) Symmetrical distribution



The spread of the frequencies is the same on both sides of the centre point of the curve.

(b) Positively Skewed distribution



The value of the mean is maximum and mode is least, the median lies in between the two. The distribution of frequencies are spread on the right hand than on the left.

(c) Negatively Skewed distribution

CHAPTER FIVE:

MEASURES OF DISPERSION

5.1 THE RANGE

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols Range = L- S

Where L = Largest value

S= Smallest value

COEFFICIENT OF RANGE

Coefficient of Range=
$$\frac{L-S}{L+S}$$

Example:
Find the value of range and it's co-efficient for the following data
7, 9, 6, 8, 11, 10, 4
Solution L=11 State
 $T_{11} = 11 - 4$
Coefficient of Range = $\frac{L-S}{L+S}$
 $= \frac{11-4}{11+5}$
 $= \frac{7}{15} = 0.4667$

CHAPTER SIX

REGRESSION AND CORRELATION:

6.1 Definition

Regression

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

Types of regression

Regression analysis can be classified into

 \checkmark Simple and Multiple

Considers only two variables e.g influence of advertising on sales vrnover

✓ Linear and Non Linear

while Multiple considers more than two variables **CO** Linear and Non Linear Linear are based on straighthree trend while non-linear involve curved line trends Total and Partial



Total relationship considers all the important variables, while in partial not all but a few.

LINEAR REGRESSION EQUATION

The Linear regression equation of Y on X is

Y = a + bX....(1)

And X on Y is

X = a + bY. (2)

a, b are constants.

For regression analysis of two variables there are two regression lines, namely Y on X and X on Y. The two regression lines show the average relationship between the two variables.

6.2 SCATTER DIAGRAMS

In regression, a line is drawn between these points either by free hand or scale rule in such a way that the squares of the vertical or the horizontal distances between the points and the line of regression is the least.

PRINCIPLE OF LEAST SQUARES

A line fitted by the method of least squares is known as the line of best fit. The line has the following rules

- The algebraic sum of deviation in the individual bervations with reference to the regression line may be equil to zero.
- ✓ The sum of squares of these deviations is than the sum of squares of the squar
- The lines of regression (best fit) intersect at the mean values of the variables X and Y.

6.3 LINEAR REGRESSION- LINE OF BEST FIT

Methods of Regression Analysis

- ✓ Graphic method (Scatter diagrams)
- ✓ Algebraic Method- 1. Regression equations through regression coefficient

2. Regression Equations through normal equations

ALGEBRAIC METHOD

9.4.2 Algebraic Methods:

(i) Regression Equation.

The two regression equations

Multiplying (3) by 8, we get

 $240 = 40a + 320 b \dots (5)$ (4) - (5) gives -26 = 20b $b = \frac{-26}{20}$ = -1.3Substituting b = -1.3 in equation (3) gives 30 = 5a - 52 5a = 82 $a = \frac{82}{5}$ = 16.4Hence, required regression line of X on Y is X = 16.4 - 1.3Y
6.4 CORRELATION INDCORRELATION ODEFFICIENT CORRELATION INDCORRELATION ODEFFICIENT

Correlation refers to the relationship of two variables or more. It is a statistical analysis which measures the degree or extent to which the two variables fluctuate with reference to each other.

Uses of Correlation

- \checkmark It is used in physical and social sciences
- ✓ It is useful for economists in the study of relationship between variables like price, quantity e.t.c
- ✓ Useful in measuring the degree of relationship between variables like demand and supply e.t.c
- \checkmark Sampling error can be calculated
- $\checkmark\,$ It is the basis for the concept of regression