CHAPTER ONE

DATA DESCRIPTION AND NUMERICAL MEASURES

1.0 Introduction

In this chapter, there are two data classifications such as Qualitative and Quantitative. The qualitative data can be classified into nominal and ordinal data however the quantitative data can be classified as discrete and continuous data. Graphical and numerical methods are explained for both qualitative and also quantitative. Numerical measures will cover the measures of central tendency and the measures of dispersion.

1.1 Definition

Statistics is a field of study which implies collecting, presenting, analyzing and interpreting data as a basis for explanation, description and comparison. There two types of statistics such as descriptive statistics and inferential statistics.

Descriptive statistics is a study which involves organizing, displaying and describing data by using tables, graphs and summary measures.

Inferential statistics consists of generalizing from samples to populations, performing hypothesis tests, determining relationships among variables, and making predictions.

Population refers to every element in an observation which are of interest for data collection.

Sample refers to a certain number of elements that have been chosen from a population for observation. Sample is subset to population.

Raw data is the data that have been collected before they are processed and ranked.

Frequency distribution is the classes and the number of elements or values that belongs to each of the categories or classes which can be divided into two types such as ungroup frequency (category type) and group frequency (interval type).

Class interval is a range of values defined by the lower class limit and the upper class limit.

Class boundary is the midpoint of the upper limit of one class and the lower limit of the next class.

Class midpoint is an average of lower class limit and upper class limit.

Range is equal to the highest value minus the lowest value.

Number of classes can be obtained by using Sturge's formula.
Number of classes = 1 + 3.3 log n where n : observation number in data set

Class size can be obtained by dividing the range with a number of classes.

Cumulative relative frequency can be obtained by using formula:

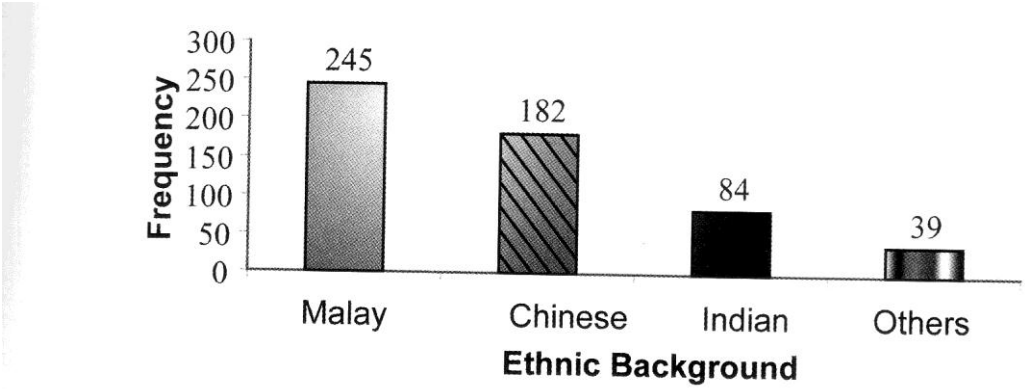
$$Cumulative\ relative\ \ frequency = \frac{cumulative\ \ frequency\ \ of\ each\ class}{sum\ of\ all\ frequencies}$$

1.2 Graphing Grouped Data

After the data have been organized into a frequency distribution, they can be represented in graphic forms such as histograms, frequency polygon and ogives.

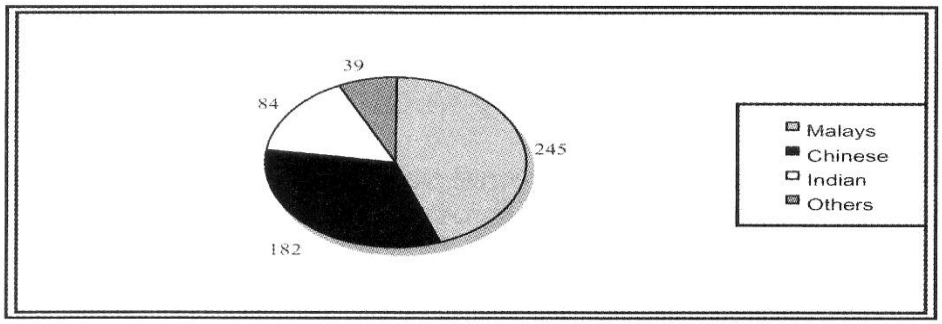
Some examples of charts

i) Bar chart

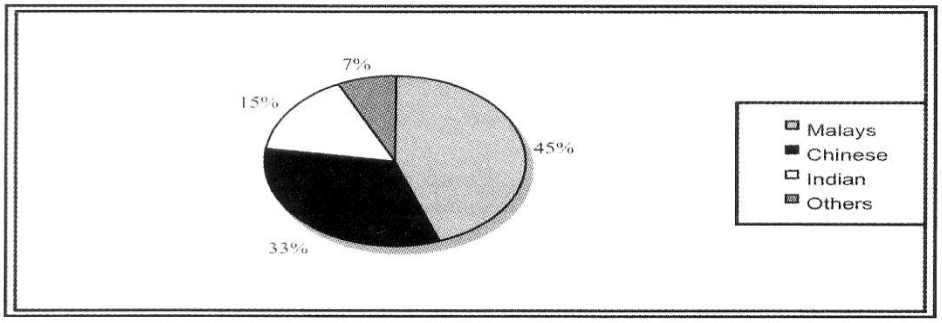


Bar chart for the number of students by their ethnic background in School J.

ii) Pie chart

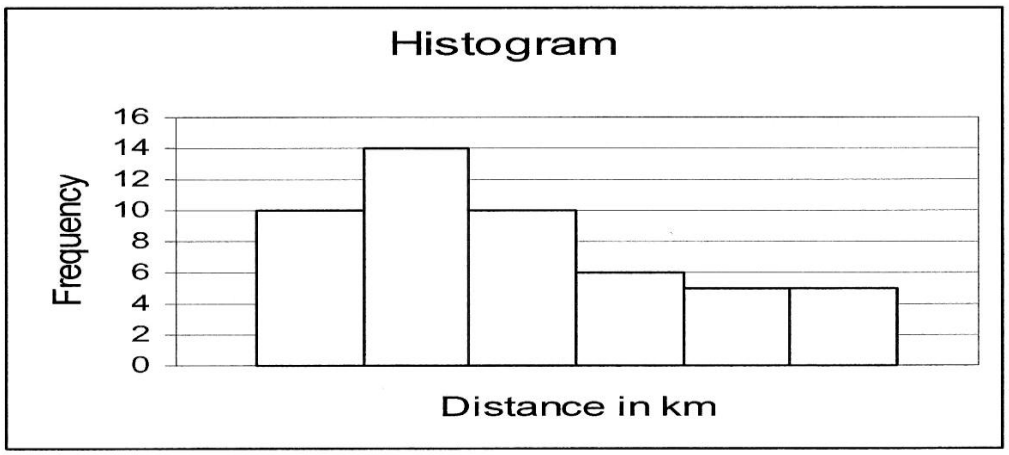


Using frequency to develop the Pie Chart

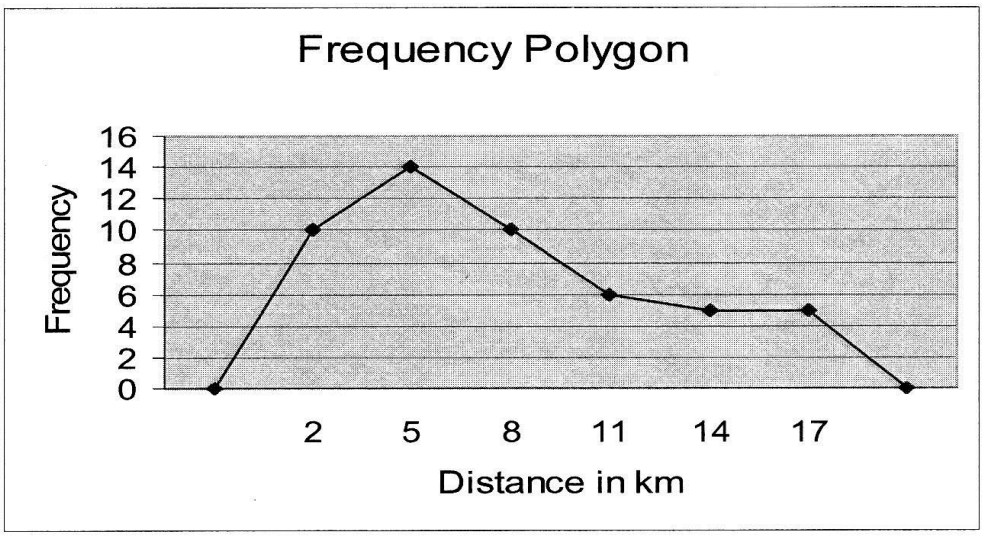


Using Relative Frequency to develop the Pie Chart

iii) **Histrogram**



iv) Frequency Polygon



1.3 Measures of central tendency

Three common measures of central tendency are mean, median and mode.

Mean, $\bar{x}$ is the average from sample and mean, μ is the average from population.

i) For individual data,

$$\bar{x} = \frac{\sum x}{n} \qquad \text{and} \qquad \mu = \frac{\sum x}{N}$$

ii) For ungrouped frequency,

$$\bar{x} = \frac{\sum fx}{f}$$ where f is frequency

iii) For grouped frequency,

$$\bar{x} = \frac{\sum fm}{f}$$ where m is midpoint

Median is the value of the item which is located at the center of the distribution.

For individual and group data, the location of median = $\frac{n+1}{2}$ th term

Mode is the value which occurs most frequently in a distribution

Range is the difference between the highest and the lowest value.

1.4 Variance and standard deviation

The standard deviation measures the spread of the data as compared to the mean.

i) Individual data

Standard deviation, $\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$ for sample distribution

Standard deviation, $s = \sqrt{\frac{\sum x^2}{n-1} - \frac{\left(\sum x\right)^2}{n(n-1)}}$ for population distribution

ii) Ungrouped frequency

Standard deviation, $\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$ for sample distribution

Standard deviation, $s = \sqrt{\frac{\sum fx^2}{\sum f - 1} - \frac{\left(\sum fx\right)^2}{\sum f \sum (f-1)}}$ for population distribution

iii) Grouped frequency

Standard deviation, $\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left(\frac{\sum mf}{\sum f}\right)^2}$ for sample distribution

Standard deviation, $s = \sqrt{\frac{\sum fm^2}{\sum f - 1} - \frac{\left(\sum fm\right)^2}{\sum f \sum (f-1)}}$ for population distribution

where m is the midpoint

1.5 Variable

A **variable** is an interested criterion to be measured on each individual such as height, or *weight or a criterion to be observed.*

The value of **Discrete Quantitative Variabale** is integer by nature which is obtained by counting process.

The value of **Continuous Quantitative Variabale** is obtained through a measuring process.