

INTRODUCTION TO STATISTICS

MATH0102

Prepared by: Nurazrin Jupri

STATISTICS ???

- * Numerical facts –
eg. the number of people living in a certain town, or the number of cars using a traffic route each day
- * The study of ways of collecting and interpreting these facts

Why We Needs to Know about Statistics

- * To know how to properly present information
- * To know how to draw conclusions about populations based on sample information
- * To know how to obtain reliable forecasts

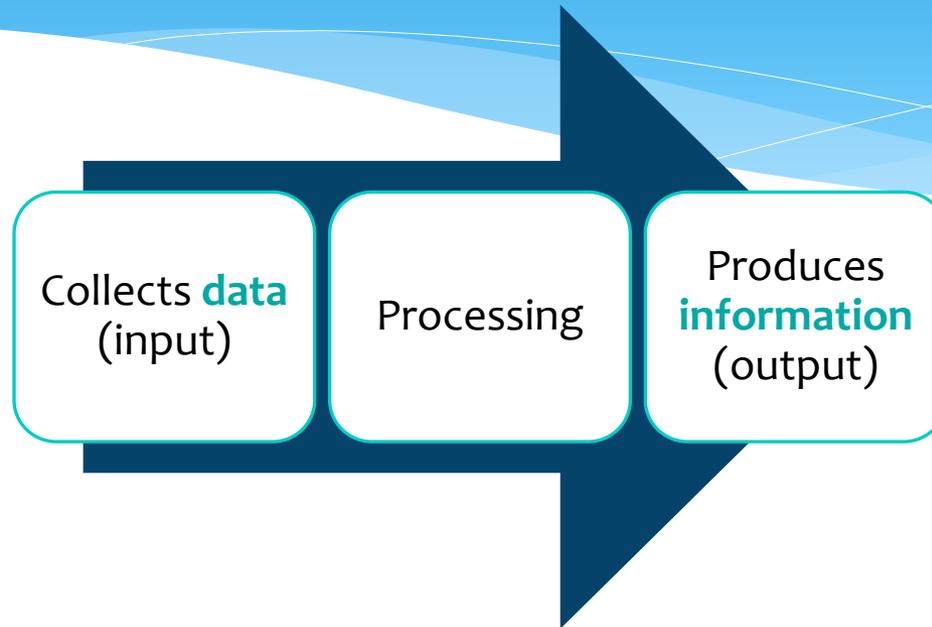
Types of Statistics

Descriptive Statistics	Inferential Statistics
The area of statistics concerned with organizing and summarizing the inevitable variability in collections of actual observations or scores.	The area of statistics that provides tools for generalizing beyond collections of actual observations
Tables, graphs, averages, ranges, correlations	Permits us to use a relatively small collection of actual observations to evaluate

Test your understanding...

No	Statement	Descriptive	Inferential
1	A graph showing the annual change in global temperature during last 30 years		
2	A graph shows that 70% of students are female		
3	In Malaysia, a researcher's hypothesis that there a positive relationship between gender and course selections for higher education study.		

Data vs Information

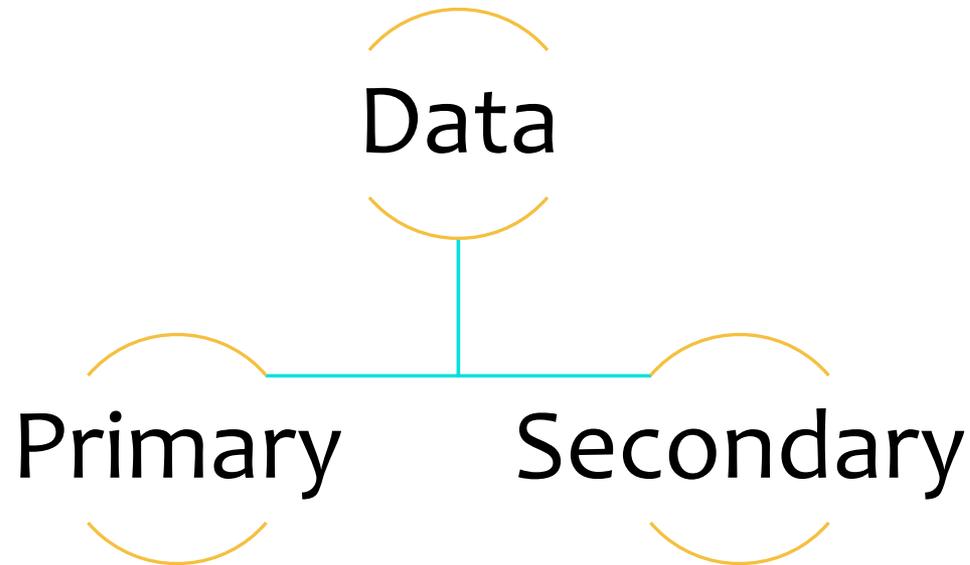


Data are the raw materials for data processing. **Information** is data that has been processed

Collecting Data

- * Data can be collected in a number of ways
Eg: monthly reports from all their departments on such matters as sales, purchases, capital expenditure, wages paid, staff employed, and so on.
- * Can then be used to **compare actual performances with planned performances**, and **adjustments** may be made in order to achieve better results.

Types of data



Primary data

Data **collected** especially for the purpose of whatever survey is being conducted.

Raw data are primary data which have not been processed at all, and which are still just a list of numbers.

The main sources of primary data are **personal investigation, teams of investigators, interviews, questionnaires and telephone surveys.**

It is **reliable** as you know where the data has come from and are aware of any inadequacies or limitations.

However, it can **take time** to collect and is **expensive**.

Secondary data

Data which have **already been collected** elsewhere, for some other purpose, but which can be used or adapted for the survey being conducted.

For example from **government, banks, newspapers, the Internet.**

Secondary data sources may be satisfactory in certain situations, or they may be the only convenient means of obtaining an item of data.

It is essential to ensure secondary data used is accurate and reliable.

relatively **inexpensive** and **quick** to obtain

Types of variables

Variable

The characteristics of the population of interest

Quantitative / Numerical

Measured on numerical scale
Yield numerical response

Qualitative / Attribute

Measured on non-numerical scale
Yields categorical response

Discrete

Arises from a
counting process

Continuous

Arises from a
measuring process

Discrete data

- * Can only be whole numbers,
- * e.g. people, houses



Continuous data

Can take on *all* values: examples include

- * Change in speed (acceleration) of a car
- * Change in height of a person
- * Change in age of a person



Quantitative data

Discrete data

are the number of goals scored by Arsenal against Chelsea in the FA Cup Final:

Arsenal could score 0, 1, 2, 3 or even 4 goals (**discrete variables** = 0, 1, 2, 3, 4), but they cannot score 2.1 or 2.5 goals.

Continuous data

include the heights of all the members of your family, as these can take on any

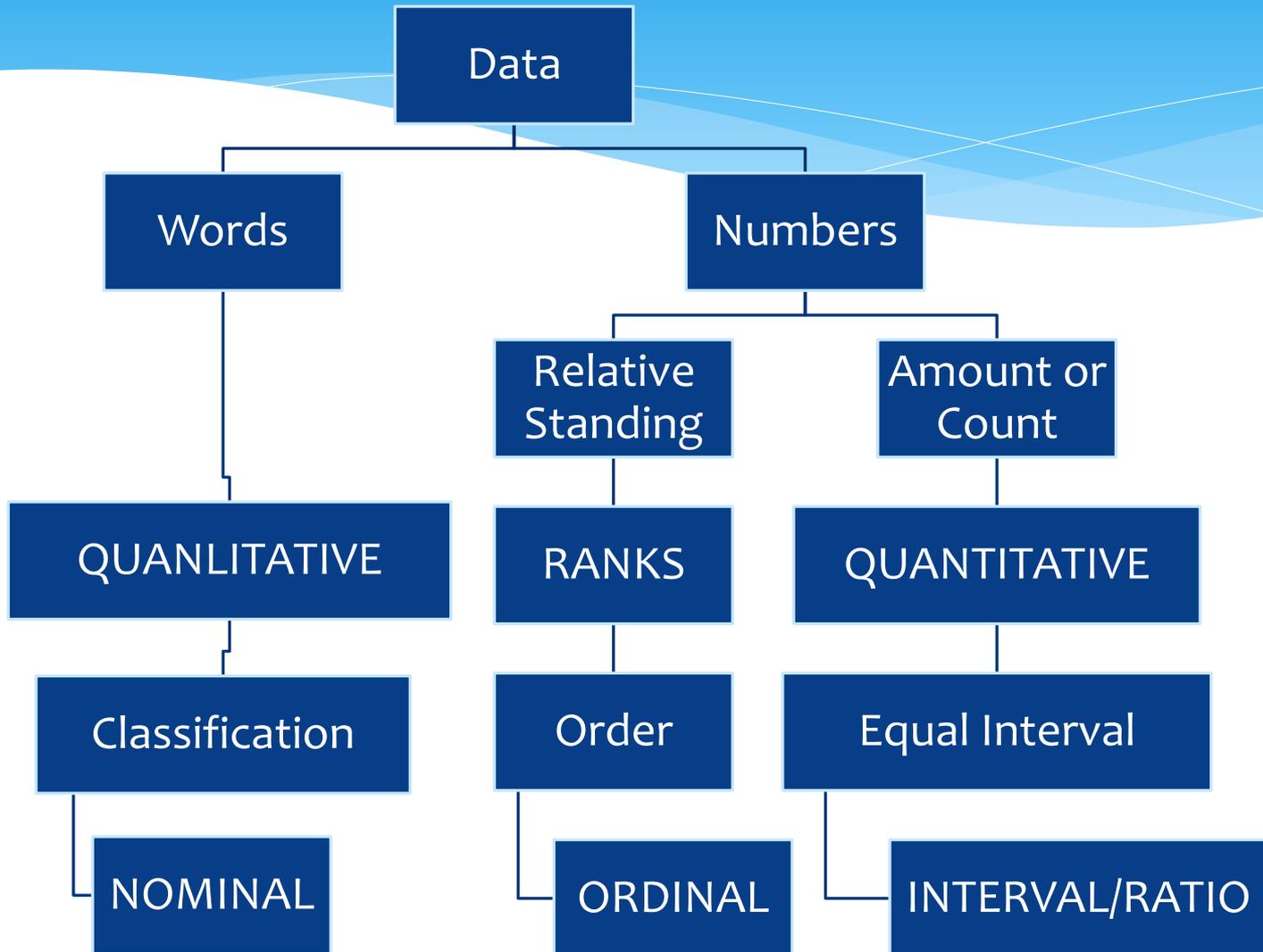
value: 1.542m, 1.639m and 1.492m for example.

Continuous variables = 1.542, 1.639, 1.492

Look through the following list of surveys and decide whether each is collecting qualitative data or quantitative data. If you think the data is quantitative, indicate whether it is discrete or continuous.

Statement	Qualitative	Discrete	Continuous
(a) A survey of accountancy textbooks, to determine how many diagrams they contain.			
(b) A survey of greetings cards on a newsagent's shelf, to determine whether or not each has a price sticker on it.			
(d) A survey of swimmers to find out how long they take to swim a kilometre.			

Scale of Measurement



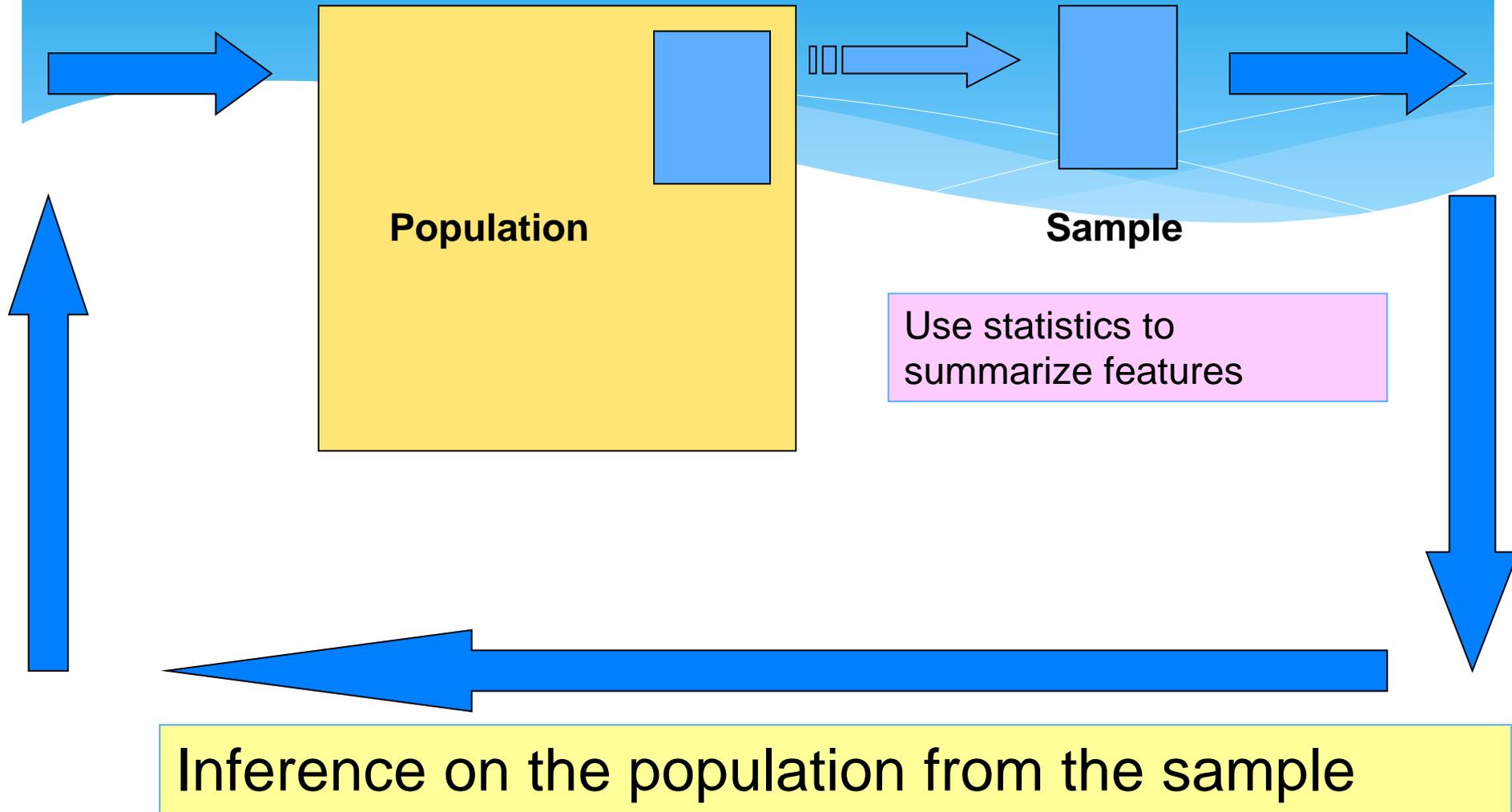
Population vs Sample

- * A **population** (universe) is the collection of things under consideration
- * A **sample** is a portion of the population selected for analysis

Data are often collected from a **sample** rather than from a population.

If the whole population is examined, the survey is called a **census**.

Population and Sample



Simple Random

A sampling frame is constructed . So every item has an equal chance of being selected

Multistage

Divides population into sub populations then selects sample of sub-populations at random

Stratified Random

Population is divided into strata or categories and random samples taken from each strata

Sampling methods

Cluster

Non-random, selects **one definable subsection** of population as the sample

Quota

Interviewers interview all the people they meet up to a certain quota

Fig. 11.4

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

A Graphical Illustration of Simple Random Sampling

Select five random numbers from 1 to 25. The resulting sample consists of population elements 3, 7, 9, 16, and 24. Note, there is no element from Group C.

Fig. 11.4

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

A Graphical Illustration of Stratified Sampling

Randomly select a number from 1 to 5 for each stratum, A to E. The resulting sample consists of population elements 4, 7, 13, 19 and 21. Note, one element is selected from each column.