

COMM121 – Statistics for Business

CHAPTER 1: Introduction and Data Collection

Statistics: is concerned with processing and analysing data, and collecting, presenting and transforming data to assist decision-makers.

Key definitions:

Population: all the members of a group about which you want to draw a conclusion.

Sample: the portion of the population selected for analysis.

Parameter: a numerical measure that describes a characteristic of a population.

Statistic: a numerical measure that describes a characteristic of a sample.

Descriptive statistics: collecting, summarising, presenting, and characterising data.

Important sources of data:

- Data distributed by organisation or individual
- Designed experiment
- Survey
- Observational study

Data sources can be classified as:

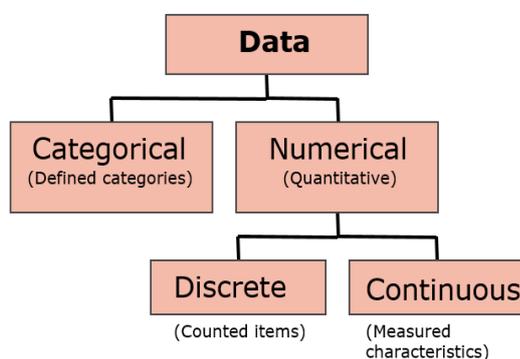
- Primary sources (when the data collector is the one using the data for analysis).
- Secondary sources (when another organisation or individual has collected the data that are used for analysis by an organisation or individual).

Inferential statistics: drawing conclusions about a population based on sample data (i.e. estimating a parameter based on a statistic).

Estimation: e.g. estimate the population mean weight (parameter) using the sample mean weight (statistic).

Hypothesis testing: e.g. test the claim that the population mean weight is 100 kilos.

Types of data:

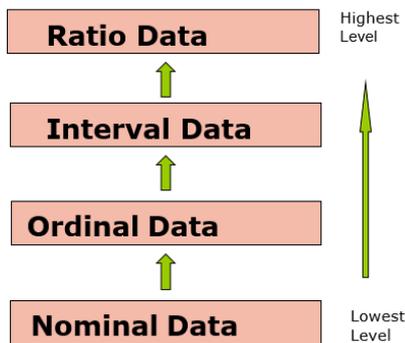


Categorical: simply classifies data into categories (e.g. marital status, hair colour, gender).

Numerical (discrete): counted items – finite number of items (e.g. number of children, number of people who have type-O blood).

Numerical (continuous): measured characteristics – infinite number of items (e.g. weight, height).

Levels of measurement and measurement scales:



Nominal: categorical data with no ranking or direction, e.g. colours, animal species, and gender.

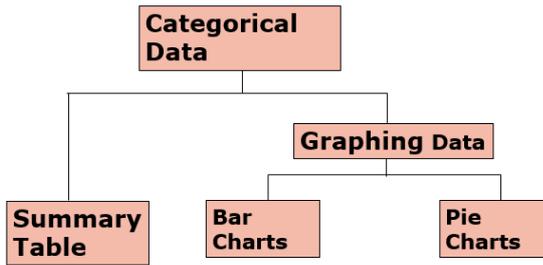
Ordinal: categorical data with ranking, e.g. education, age categories, and student letter grades.

Interval: continuous numerical data where differences are meaningful but there is no true zero, e.g. temperature and year.

Ratio: continuous numerical data where differences are meaningful and there is a true zero point, e.g. time engaged on a task, height, weight, age, weekly food spending.

CHAPTER 2: Presenting Data in Tables and Charts

Tables and charts for categorical data:



Summary table:

| Investment Type | Amount (in 000s \$) | Percentage (%) |
|-----------------|---------------------|----------------|
| Stocks | 46.5 | 42.27 |
| Bonds | 32.0 | 29.09 |
| CD | 15.5 | 14.09 |
| Savings | 16.0 | 14.55 |
| Total | 110.0 | 100.0 |

Bar charts and pie charts: are often used for qualitative data (categories or nominal scale). The length of the bar or size of pie slice shows the frequency or percentage for each category.

1. **Bar charts:** are preferred for comparing categories.
2. **Pie charts:** are preferred for observing the portion of the total (e.g. market share).

