STATISTICAL INVESTIGATIONS – TYPES OF DATA

Introduction video to descriptive statistics

https://www.youtube.com/watch?v=eIZD1BFfw8E

The Australian Bureau of Statistics (ABS) collects information about Australians and their daily lives. The largest and one of the most important ways they do this is through the Census of population and Housing, which is conducted every five years; it is a descriptive count of everyone who is in Australia on one particular night, what they do and where they live.

Because a **census** includes every member of a target population, it is a huge and costly undertaking, which is why the ABS Census is only conducted every five years. Surveys that involve **samples** of the whole population are used more frequently to predict trends for the whole population.

Conducting a statistical investigation

A statistical investigation should typically include four phases:

Phase 1: Identify the problem and set a statistical question

Phase 2: Collect statistical data

Phase 3: Analyse the data

Phase 4: Interpret and communicate the results.

To help understand these phases, consider how the phases apply in a simple example. Imagine that you are conducting a small investigation into some of your classmates' characteristics. You will be collecting information on the following:

- gender
- left-handed or right-handed
- arm span (to nearest centimetre)
- position in family (eldest, middle or youngest child).

Phase 1: Identify the problem and set a statistical question.

The problem is to discover whether connections exist between pairs of these characteristics for the students in your class. For example, you could compare students based on whether they are right-handed or left-handed and their position in the family.

Phase 2: Collect the statistical data.

As a data collection device you could use a simple card as shown.

GENDER: FEMALE	ARM SPAN: 153 cm Stretch out your arms and measure from finger tip to finger tip
LEFT-HANDED RIGHT-HANDED Circle the correct response	YOUNGEST MIDDLE ELDEST Circle the correct response

Phase 3: Analyse the data.

After you have collected data, to analyse it you need to display the data. This can be done in a table, a bar graph, a dot plot, a stem-and-leaf plot or a histogram. (These display tools will be discussed in more detail later.) Two-way tables could be used to pair up the variables as shown below. Note that e.g. 150–<160 means values including 150 but less than 160.

Arm span (cm)	Male	Female
150–<160	2	4
160–<170	3	5
170–<180	3	2
180–<190	2	1
190–<200	2	0

Place in family	Left-handed	Right-handed
Youngest	3	8
Middle	2	4
Eldest	1	6

Phase 4: Interpret and communicate the results.

For the final part of the process you will need to write some conclusions about the data. Some examples based on the sample data recorded above could be:

- The median arm span for males was greater than the median arm span for females.
- A greater percentage of right-handed students were the eldest in the family compared with left-handed students.

You could also draw some appropriate statistical graphs to help explain the connections you have found.

Categorical and numerical data

Data is information that will most often be raw facts that can be collected and/or measured. There are two basic types of data:

- **Categorical data** can be grouped into categories. Examples include hair colour and position in the family.
- **Numerical data** has values that can be counted or measured. Examples include height and number of brothers and sisters.

There are two types of categorical data:

- **Nominal data** has no special order associated with the possible responses. Examples include hair colour.
- **Ordinal data** does have some sort of order associated with it. Examples include position in family.

There are two types of numerical data:

- **Discrete data** has a countable number of numerical values, but they don't have to be whole numbers. Examples include the number of brothers and sisters, or the age of your teacher.
- **Continuous data** has an infinite number of possibilities within a particular range. This type of data is measured using some sort of instrument. The accuracy with which we record the data may vary but does not alter the fact that there are an infinite number of possible values. Examples include the heights of a group of people.