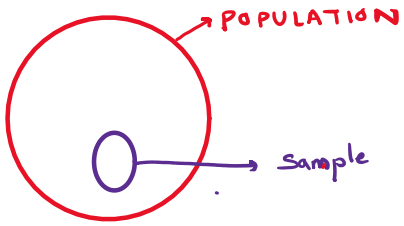
Week 1

Monday, 23 October 2023 9:51 AM

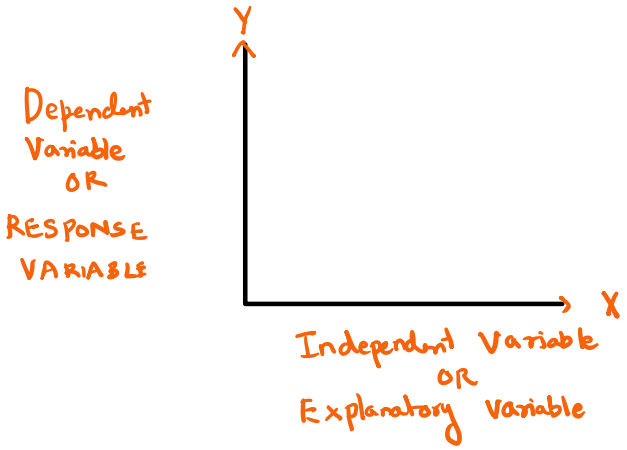
- **Population:** a collection of objects whose properties are to be analysed
  e.g. all students attending Deakin
- **Sample:** a subset or part of a population
  e.g. a random selection of Deakin students

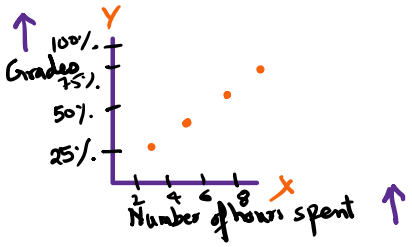


Example : Population → All deakin college students

Sample → students taken SIT191 in T3 2023

- **Explanatory variable:** a variable that is being manipulated by the researcher
- **Response variable:** a variable whose values are to be compared between treatment groups





- **Parameter:** a numerical characteristic that describes a population
  e.g. average age of all students

- **Summary statistic:** a numerical characteristic that describes a sample
  e.g. average age of a sample of Deakin students

population parameters

→ describes population

→ Greek symbols

Mean or Average → μ

Standard Deviation → σ

Sample size → N

Sample statistics

describes Sample

→ Latin symbols

Mean or Average → $\bar{y}$

Standard deviation → s

Sample size → n

Question 1 : Identifying sample statistics, population parameters and the sample size (*n*)

1. A survey of randomly selected 18-24 year olds reports that 83% watch a movie at least once a month.
a) Identify the population of interest. 18-24 year olds
b) Is the 83% a parameter or a statistic? 83% → Sample statistic

2. 11% of a random sample of 100 people have a particular gene. It is thought that 15% of all people have the gene.
a) Is 11% a statistic or parameter? Sample statistic
b) Is 15% a statistic or parameter? population parameter
c) Give the sample size, *n*. n=100

3. Suppose the average length for a species of worm is 6.3mm. A random sample of 30 such worms had a mean length of 7.8mm.
a) Is 6.3mm a parameter or a statistic? population parameter
b) Is 7.8mm a parameter or a statistic? Sample statistic
c) What is *n*? n = 30

Types of variables

Categorical variables QUALITATIVE VARIABLE

Examples: gender, ethnicity, favourite football team, postcode
'Categorical' (or *qualitative*) refers to naming or classifying things. If words are the data values, then the variable is categorical.
An ***ordinal categorical variable*** is one where data can be compared as higher or lower on a given scale. There is a natural order to the values, such as "Low", "Medium" and "High".
Examples include a Four-point Likert Scale (Strongly Disagree, Disagree, Agree, Strongly Agree) or your final grade (HD, D, C, P, N).
For a ***nominal categorical variable*** no sensible order to the value exists. A student's usual mode of transport to school, hair colour or favourite football team are examples of nominal data.
It is possible for some numerical values to be categorical data if the values have an identifying or labelling function only. For example, mobile phone numbers, post codes or football players' numbers act to identify rather than measure something and so would be classed as categorical variables. Such information won't have units, and it would not make sense to calculate the average.

Numerical variables QUANTITATIVE NUMBERS

Examples: weekly income ($), height (cm), weight (kgs), no. of students enrolled
Numerical (or *quantitative*) variables have data that are numbers with numerical meaning, and units. Examples include people's height and weight, the number of children in a family and the distance travelled by students to school.

Discrete numerical variables have no values possible in between successive values, or can be thought of as having a finite number of values between any two given values. For example, the number of pets in a family, or the number of subjects you are enrolled in are discrete variables. The data are often whole numbers but there are exceptions. For example, shoe sizes (6, 6½, 7, 7½, 8 etc.) are not always whole numbers, but there are no values in between successive sizes.
Number of units taken
Continuous numerical variables have, in theory, an infinite number of values possible between any two given values. For example, height and weight are continuous variables, since people are not exactly 165 cm or 178 cm, but could be 165.3 cm or 178.2034 cm. In practice though, measurements are restricted to a certain level of accuracy and so may not *appear* to be continuous.

Question 2 : Identifying the variable type

Identify the variable type for the following as continuous or discrete numerical, or nominal or ordinal categorical.

1. Fish species Nominal Categorical
2. Fish length Continuous Numerical
3. No. of eggs laid by birds Discrete Numerical
4. Course students are enrolled in Nominal Categorical
5. Your final grade (HD, D, C, etc) for a subject. Ordinal categorical
6. Tax file number Nominal Categorical

The "W's"

- To provide context of a study or experiment, we need the 5 W's and the How of the data.
  - Who
  - What (and in what units)
  - When
  - Where
  - Why
  - and How
- Note: sometimes it is not possible to determine all of these

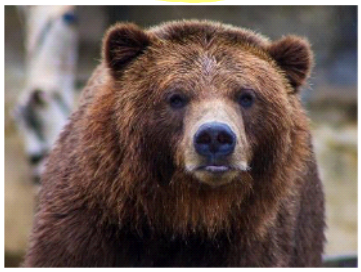
?????

???

Question 3 : Identifying the 5Ws and other terms

Because of the difficulty of weighing a bear in the woods, researchers caught and measured 54 bears, recording their weight, neck size, length, paw print area and depth, and gender. They hoped to find a way to estimate weight from the other more easily determined quantities.

Why



1. Identify the 5Ws and How (if possible).
2. Identify the target population and the sample.
3. Is the average weight of the sampled 54 bears a statistic or a parameter?
4. How would the data matrix for this study be structured?

- Who → Bears / Bear in the woods
- What (and in what units)
- When → not stated / not given
- Where → not stated / in the woods
- Why
- and How → not stated

Categorical
Gender

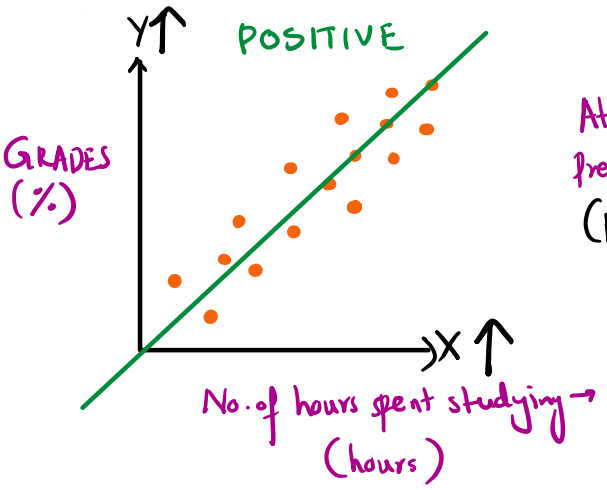
Numerical
Weight (kg)
Neck size (m)
Length (m)

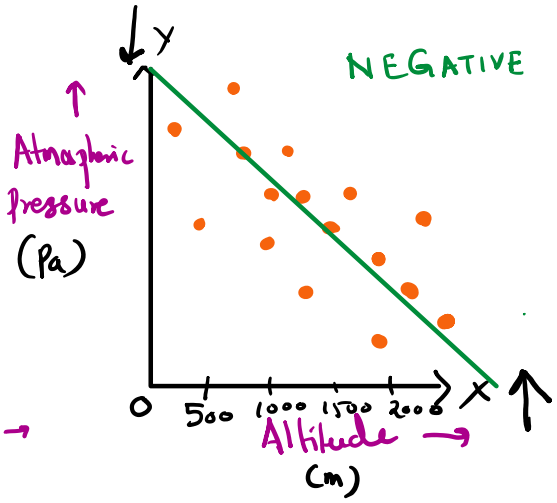
- Why
- and How → not stated

Length (m)
paw print area (m^2)
depth (m)

Scatterplots are a type of graph used to explore a potential relationship between two *numerical* variables.

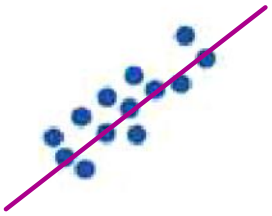
DIRECTION :

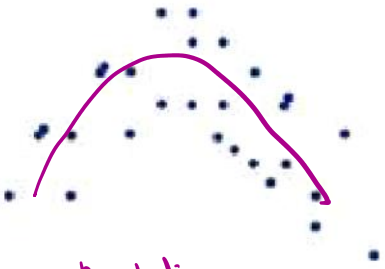




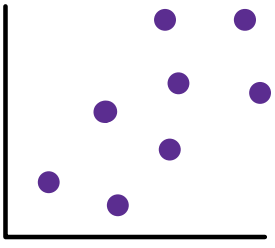
FORM:

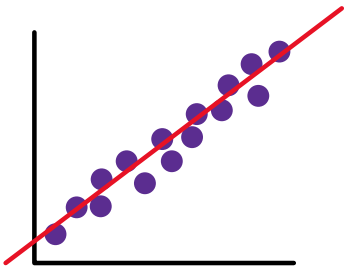
Form:
Relationships can be linear or curved.

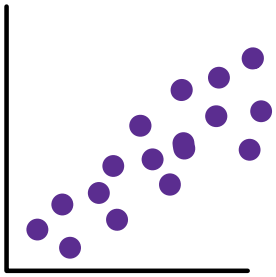




STRENGTH : Weak, Moderate, Strong







Scatterplots Which of these scatterplots show

a) little or no association? 4
b) a negative association? 3, 4
c) a linear association? 2, 3, 4
d) a moderately strong association? 2
e) a very strong association? 1, 3