

# Stats

Data = information.

Variable = characteristic of interest.  
What do we want to know about?

Observation = value given to a variable (days missed)

Population = full set of data under consideration. [6<sup>th</sup> years]

Census = collection of data from all population.

Sample = part of population. - can create bias.

Survey of sample  $\rightarrow$  census of population

$\rightarrow$  Personal  $\rightarrow$  <sup>Adv</sup> higher response <sup>Disadv</sup> Cost / Time

$\rightarrow$  Online  $\rightarrow$  Quick / cheap Response rate

$\rightarrow$  Phone  $\rightarrow$  Higher response Cost / Time

$\rightarrow$  Postal  $\rightarrow$  Large volumes Response rate

Sampling frame = list of every item in the population.

Factors 6<sup>th</sup> to get H 1

- Present
- Work
- Behaviour

Stats → collection  
→ presentation  
→ analysis  
→ interpret data.

Primary data = data I collect myself.

Secondary data = data from another source.

Type of data.

Numerical (Numbers) → discrete = single values (Family size)

→ Continuous = scale (Height) - Age

Categorical (Not numbers) → Ordinal = order (grades)  
→ Unordinal = no order (gender)

Sample = how to split the population.

Simple random sample = equal chance

Everyone in population has equal chance of being picked. Picked out of a hat.

Stratified = groups - proportion to group.  
Divide population into natural groups (age / gender) pick from each group in proportion to population.

Systematic = every n<sup>th</sup> person.  
Pick every 10<sup>th</sup> person on line register.

Cluster = groups (strata) - geography  
Put into groups and pick from an area.

Quota = certain amount  
Make strata pick same amount from each group.