JAMES COOK UNIVERSITY AUSTRALIA

Basic Statistics

Collecting Data



Collecting Data

Learning Intentions



Today we will understand:

- What is statistics?
- What is data?
- How is data gathered?
- How do we ensure data is accurate and reliable?
- Is the data representative of the population from which it was drawn?

Image accessed: http://intouchacquisitions.co.uk/in-touch-acquisitions-review-the-importance-of-business-statistics/

What is Statistics?

Statistics is the study of how to collect, organise, analyse and interpret information

- Statistics is a tool for converting data into information

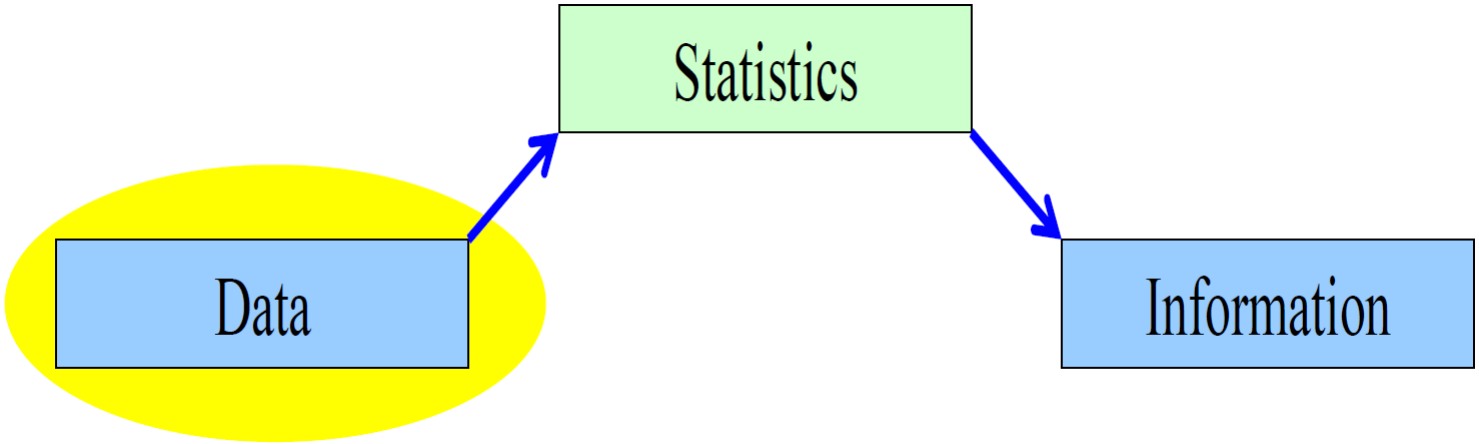


Image accessed: http://www.utdallas.edu/~scniu/OPRE-6301/documents/Data_Collection_and_Sampling.pdf

Why Study Statistics?

JAMES COOK UNIVERSITY AUSTRALIA

- Numerical information is everywhere!
- Statistical techniques are used to inform decisions that affect our everyday lives
- A knowledge of statistical methods will help you understand how decisions are made and how they might affect you
- An understanding of data analysis is helpful in most occupations



Image accessed: http://blog.businesssocialmediasolutions.com/geo-search-twitter/

Activity

Field	Example of data collected	Data used for public or private purposes?
Population		
Education		
Labour market		
Domestic Trade		
Housing market		
Medical care		
Public health		
Agriculture		
Natural Resource Management		
Welfare Services		
Law Enforcement		

What is Data?

- Data are the raw information from which statistics are created
- In reverse, statistics provide an interpretation and summary of data

Questions (what we want to know) drive the collection of data



- If you want to understand a phenomenon, you need data
- Raw data is collected as a part of research, observations and surveys

Image accessed: http://www.istockphoto.com/illustrations/data+collection#33f4eb2

Types of Data

JAMES COOK UNIVERSITY AUSTRALIA

Quantitative Data
Numerical

- Measures of values or counts and are expressed as numbers
- Relates to quantity of something, “how many” or “how much”
- QUANTITATIVE think QUANTITY

Qualitative Data
Categorical

- Measures of types and may be represented by a name, symbol or number code
- Relates to quality of something, “what type” or “which category”
- QUALITATIVE think QUALITY

Quantitative Data

JAMES COOK UNIVERSITY AUSTRALIA

Price of Car

Car Number	Price
1	$ 15 000
2	$ 23 000
3	$ 17 000
4	$ 18 000
5	$ 20 000
6	$ 25 000
7	$ 12 000
8	$ 25 000
9	$ 15 000
10	$ 23 000



Image accessed: http://cldefelice.blogspot.com.au/2009/06/avant-garde.html

Quantitative Data



- Based on a count from a distinct set of whole values
- Outcomes that can be counted and listed
- Number of heads in 100 coin tosses



- Represents measurements
- Possible values cannot be counted
- Described using intervals on the number line
- Distance from home to university



Image accessed: http://running.competitor.com/2012/06/olympic-games/out-there-flip-a-coin_54775

http://www.illustrationsource.com/stock/image/3326/a-man-taking-a-measurement-of-number-seven/?&results_per_page=1&detail=TRUE&page=9

Qualitative Data

JAMES COOK UNIVERSITY AUSTRALIA

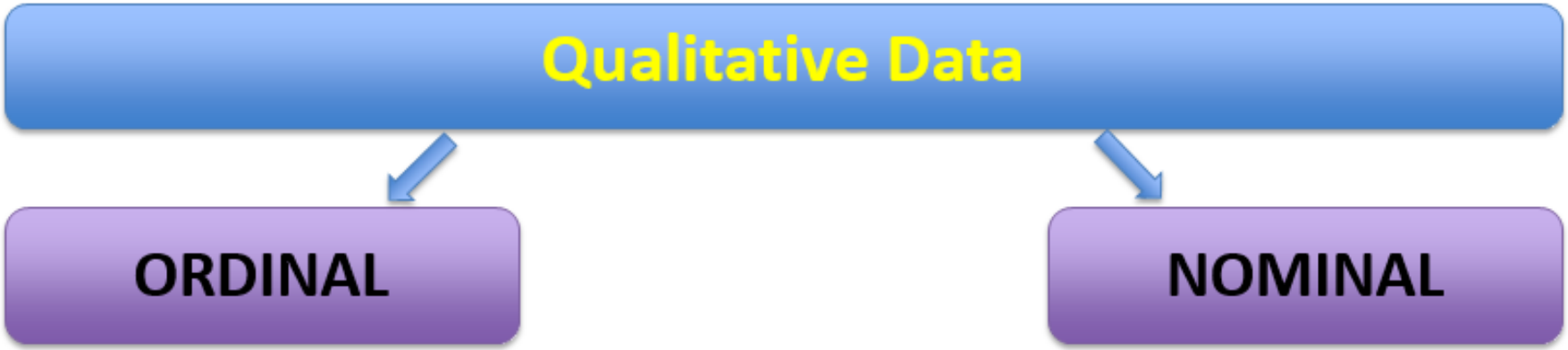
Colour of Car

Car Number	Colour
1	red
2	green
3	red
4	black
5	blue
6	green
7	black
8	blue
9	red
10	black



Image accessed: http://www.clipartpanda.com/categories/pile-20clipart

Qualitative Data

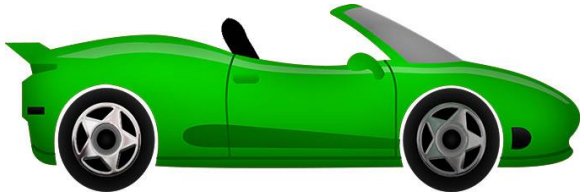


- Categories can be ordered/ranked
- Size (small, medium, large) and attitudes (strongly disagree, disagree, neutral, agree, strongly agree)
- Distance between categories can not be measured

- Categories cannot be ordered/ranked
- Gender, colour, sport

Data Unit

- A data unit is one entity in the population being studied, about which data are collected (ABS, 2013)



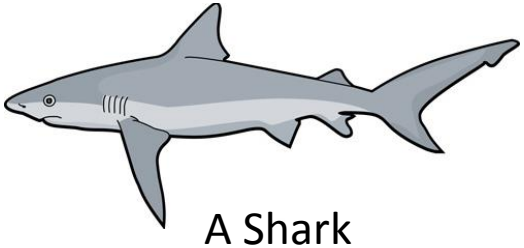
A Car



A Person



A Country



A Shark

Images accessed: http://pixshark.com/1-person-clipart.htm

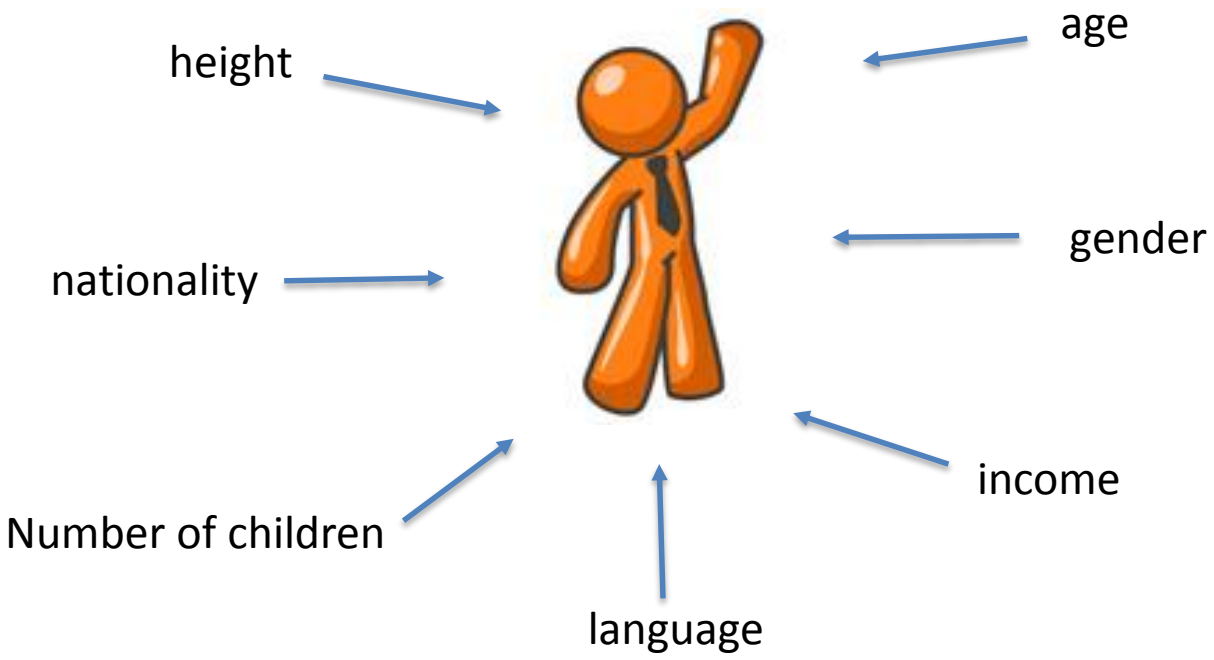
http://www.fg-a.com/autos.htm

http://www.flagsaustralia.com.au/StateFlags.html

http://classroomclipart.com/clipart-view/Clipart/Animals/Shark_Clipart/sharks_tiger_shark_728_jpg.htm

Variable

- A variable is the characteristic of the data unit being measured or counted (ABS, 2013)
- Is called a variable because the characteristic may vary between data units and may vary over time



Images accessed: http://pixshark.com/1-person-clipart.htm

Population

- A population is any complete group with at least one characteristic in common (ABS, 2013)
- It is the complete pool from which a statistical sample is drawn
- If you wanted to study the height of adult females in Australia, the population would be all adult females in Australia



- If you wanted to study the size of green ant nests on the JCU campus, the population would be all green ant nests on the JCU campus



Images accessed: http://pixshark.com/1-person-clipart.htm
http://photo.net/photodb/photo?photo_id=6801772

Sample

- Often it is not possible to measure/count every unit in a given population
- A sample is a sub-set of the population, selected to represent all units in a population of interest (ABS, 2013)
- It is a count from part of the population
- Information from the sampled units is used to infer the characteristics for the entire population of interest



A subset of the population.

Images accessed: http://www.nedarc.org/statisticalhelp/selectionAndSampling/probabilitySampling.html

A Good Sample

- Sample must be large enough to provide reliable representation of whole population

- Individuals are selected **randomly** - **each unit in the population has equal and independent chance of being selected**

- Random (or probability) sampling reduces bias and sampling error – if data is not collected randomly, it cannot be used in any meaningful way to make inferences



Image accessed: https://psychlopedia.wikispaces.com/random+sample

Simple Random Sampling

- All units of the population are chosen at random and have the same chance of being selected
- Sampled randomly throughout entire study are or study period

Question	Method
How tall are JCU students?	Assign each JCU student a number and use a random number table to select students
What is the diameter of trees on the JCU campus?	Place a grid over map of entire campus and use random number generator to select (x,y) coordinates Sample trees closest to coordinate or within quadrat

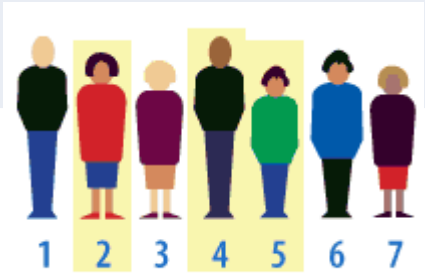


Image accessed: https://mcguiresl.wordpress.com/page/2/

Systematic Random Sample

- The first member of the sample is chosen randomly and then the others units of the sample are taken at intervals (i.e. every 5th unit)

- Appropriate when populations are distributed across zones or gradients

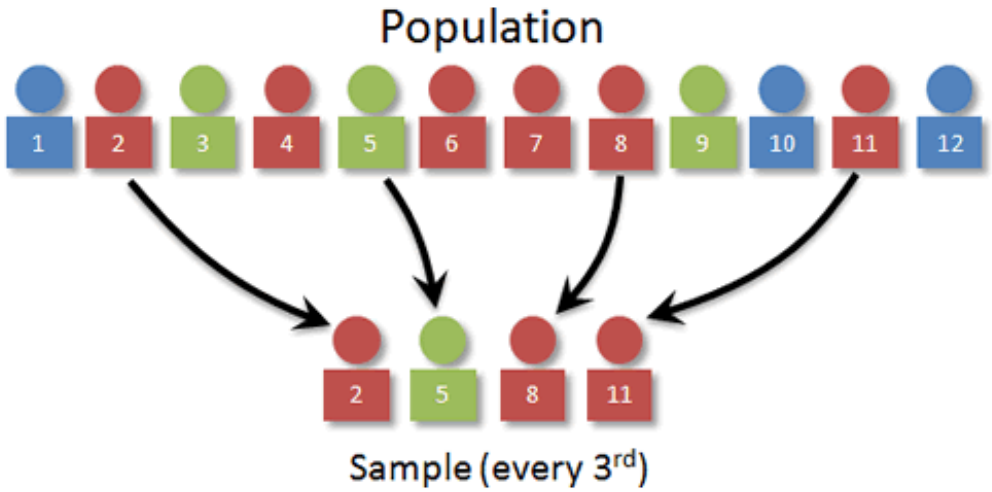


Image accessed: http://faculty.elgin.edu/dkernler/statistics/ch01/1-4.html

Stratified Random Sampling

- Relevant subgroups are identified within a population and random samples are selected from each subgroup

- Used when the population can be separated by a characteristic which may influence the variable being measured

	Example 1	Example 2
Population	All primary school students in Cairns	All people in Australia
Groups	25 different primary schools in cairns	7 states in Australia and territories
Obtain Simple Random Sample	20 students from each of the primary schools	1000 people from each state/territory
Sample	25 x 20 = 500 primary students selected	7 x 1000 = 7000

Cluster Random Sampling

- The population is divided into groups (clusters) and a simple random sample of clusters is obtained to identify a sample of clusters

- Data is obtained on every unit within each of the randomly selected clusters

	Example 1	Example 2
Population	All primary school students in Cairns	All high school basketball players in Queensland
Groups	25 different primary schools in cairns	35 different high school basketball teams in QLD
Obtain Simple Random Sample	10 primary schools randomly selected	12 teams from the 35 teams randomly selected
Sample	Every students in the 10 selected primary schools	Every player on the 12 teams selected

Non-Probability Sampling

- Should be avoided



- Volunteer samples
- Convenience sample
- Based on human decision rather than random selection
- Statistics derived from non-probability sampling cannot infer how population might behave
- Huge potential sources of bias

Image accessed: https://www.as.uky.edu/wet-research-lab-volunteer

Confounding Factors

- When factors other than the treatment influence the results – avoid!

- Zebra finches were used to study how females choose a mate, based on the males body colour

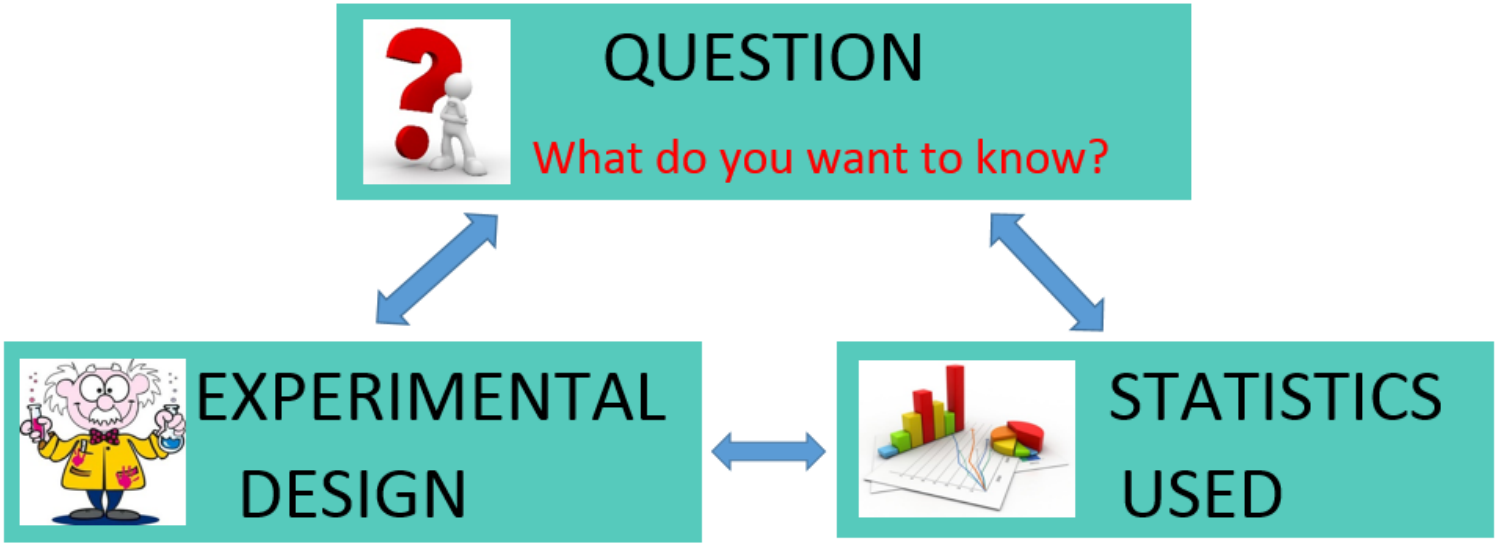
- Coloured leg bands were used to identify individuals

- Turned out that females liked certain coloured leg bands



Image accessed: https://www.as.uky.edu/wet-research-lab-volunteer

Collecting Data



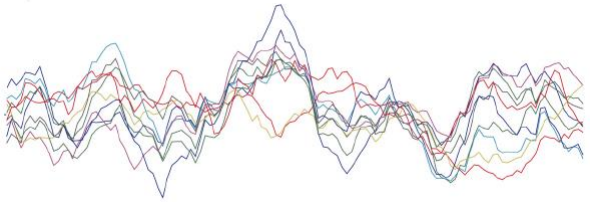
- BEFORE you collect any data, you need to know the experiment/study design and determine the statistics
- Collecting data without deciding which statistics you will use first, can result in data that cannot be analysed

Images accessed:
http://getyourthinon.com/the-scientific-method-of-nutrition/
http://hdimagelib.com/question+mark+png
http://epal-kissam.chan.sch.gr/

Collecting Data

JAMES COOK UNIVERSITY AUSTRALIA

- The world is highly variable



- Data collection is costly – in terms of money, time and resources



- It is usually not possible to measure all units in a population



- We can make inferences based on samples

Images accessed:
http://www.infolab21.lancs.ac.uk/business/isis/faq.php
http://www.exponent.com/cost_and_damages/

Collecting Data

Reflect on Learning Intentions

- What is statistics?
- What is data?
- How is data gathered?
- How do we ensure data is accurate and reliable?
- Is the data representative of the population from which it was drawn?



Image accessed: http://intouchacquisitions.co.uk/in-touch-acquisitions-review-the-importance-of-business-statistics/

References

http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language



Image accessed: http://apafolchitorres.org//