TheLearningCentre
UNLOCK YOUR POTENTIAL

JAMES COOK UNIVERSITY
AUSTRALIA

Basic Statistics Short Course

learning@jcu.edu.au

TheLearningCentre
UNLOCK YOUR POTENTIAL

Cairns
Singapore
Townsville


JAMES COOK UNIVERSITY
AUSTRALIA
STATISTICS
MEANS NEVER HAVING TO SAY YOU'RE CERTAIN
TheLearningCentre
UNLOCK YOUR POTENTIAL
Cairns
Singapore
Townsville

Today's Workshop

JAMES COOK UNIVERSITY AUSTRALIA

10 am – 12 noon

1) Introduction: Statistics and the research process
2) Types of data
3) Collecting data
4) Analysing data: Part I – Descriptive Statistics

*****12-1 pm LUNCH *****

1 pm – 3 pm

5) Data characteristics: Normal distribution
6) Analysing data: Part II – Inferential Statistics
7) Representing results

TheLearningCentre
UNLOCK YOUR POTENTIAL

3

Housekeeping

JAMES COOK UNIVERSITY AUSTRALIA

1) 

2) Sign attendance sheet

3) Fill out confidence/ability levels before workshop

TheLearningCentre
UNLOCK YOUR POTENTIAL

4

JAMES COOK UNIVERSITY AUSTRALIA

- Lecturers, tutors and supervisors
- The Learning Centre
  - https://www.jcu.edu.au/students/learning-centre/maths-and-statistics
  - Learning Advisors/Peer Advice Desk
- https://www.khanacademy.org/ KHANACADEMY
  - (good intro to chi-square test, regression, ANOVA)
- Lynda.com (structured training tutorials: e.g. Statistics foundations & applications; Excel; SPSS; R and more)
- JCU Graduate Research School 'StatsHelp' Program:
  - PhD & Research Masters
  - Professional Doctorate, or
  - Honours and coursework Masters Degree (fee basis only and pending availability).

lynda.com

TheLearningCentre
UNLOCK YOUR POTENTIAL

5

1) Statistics and the research process

JAMES COOK UNIVERSITY AUSTRALIA

What is statistics?

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

→ It is a tool for converting data into understandable *quantitative* information

TheLearningCentre
UNLOCK YOUR POTENTIAL

6

Ask students what they think statistics is.

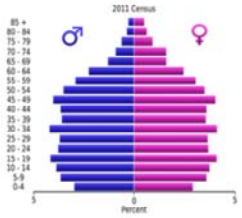
Statistics is used to make sense of raw data

1) Statistics and the research process

JAMES COOK UNIVERSITY AUSTRALIA

Why is it important?

- Numerical information is *everywhere*!
- Statistical techniques are used to *inform decisions* that affect our everyday lives
- Statistics is *part of every quantitative research process*

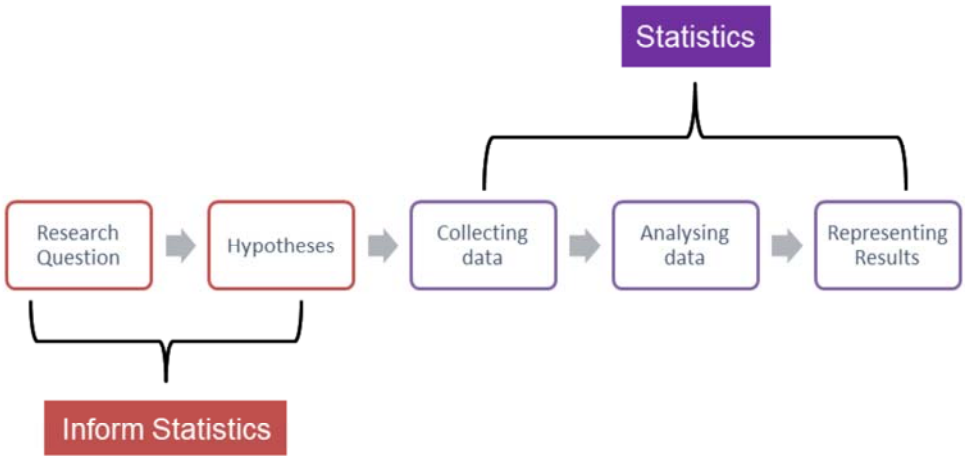


TheLearningCentre
UNLOCK YOUR POTENTIAL

1) Statistics and the research process

JAMES COOK UNIVERSITY AUSTRALIA

How does it fit into the research process?



TheLearningCentre
UNLOCK YOUR POTENTIAL

8

1) Statistics and the research process

JAMES COOK UNIVERSITY AUSTRALIA

Data? What's that?

- If you want to understand a phenomenon, you need data
- Data are collected as a part of research, observations and surveys
- Data are the *raw information* from which statistics are created
- In reverse, statistics provide an interpretation and summary of data

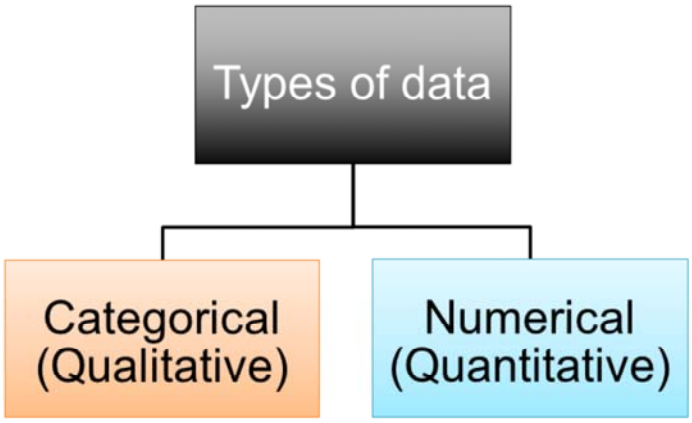
TheLearningCentre
UNLOCK YOUR POTENTIAL

9

2) Types of data

JAMES COOK UNIVERSITY AUSTRALIA

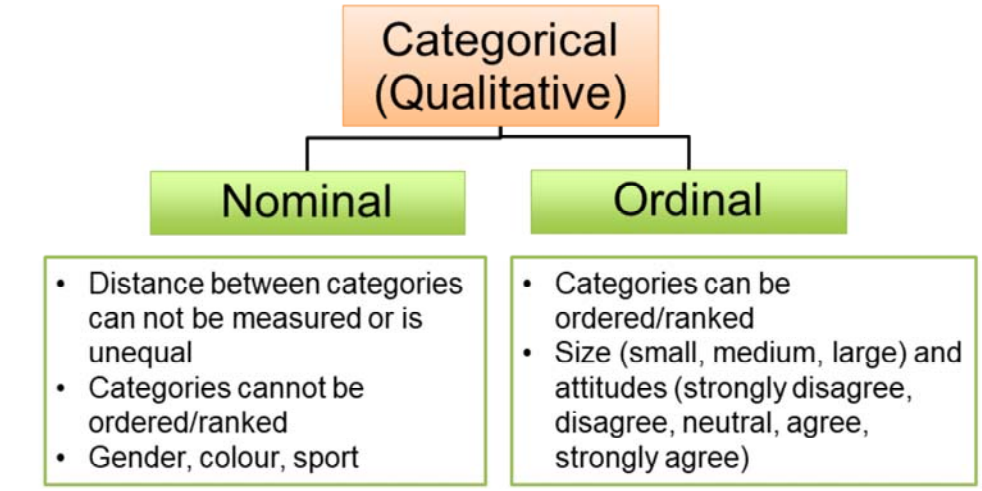
- The WHAT of collecting data



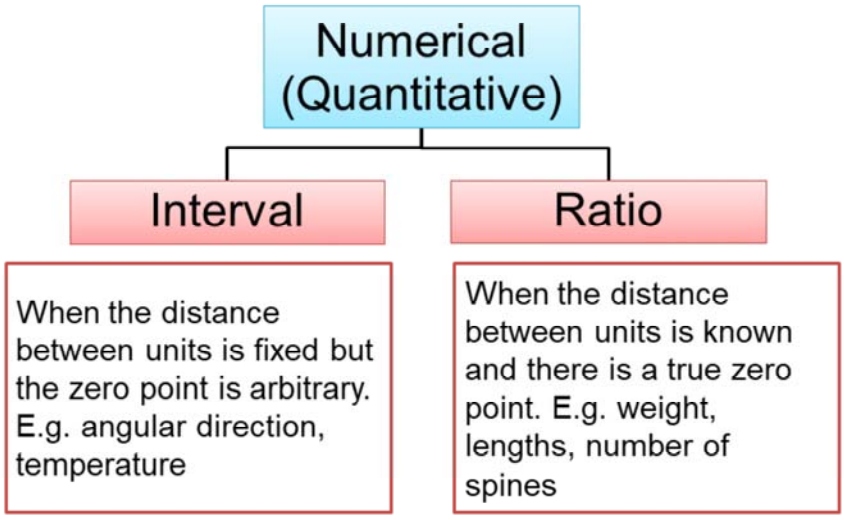
TheLearningCentre
UNLOCK YOUR POTENTIAL

10

2) Types of data

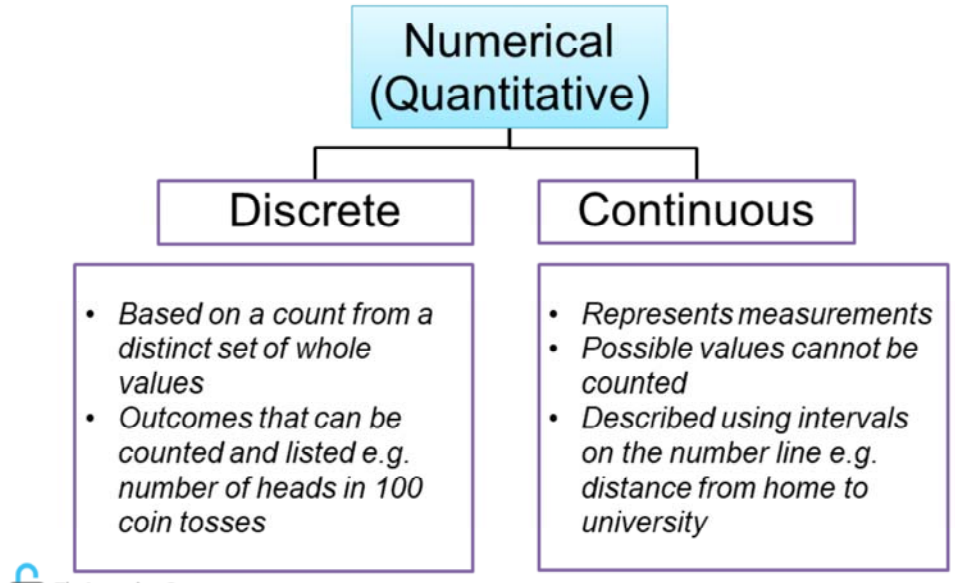


2) Types of data



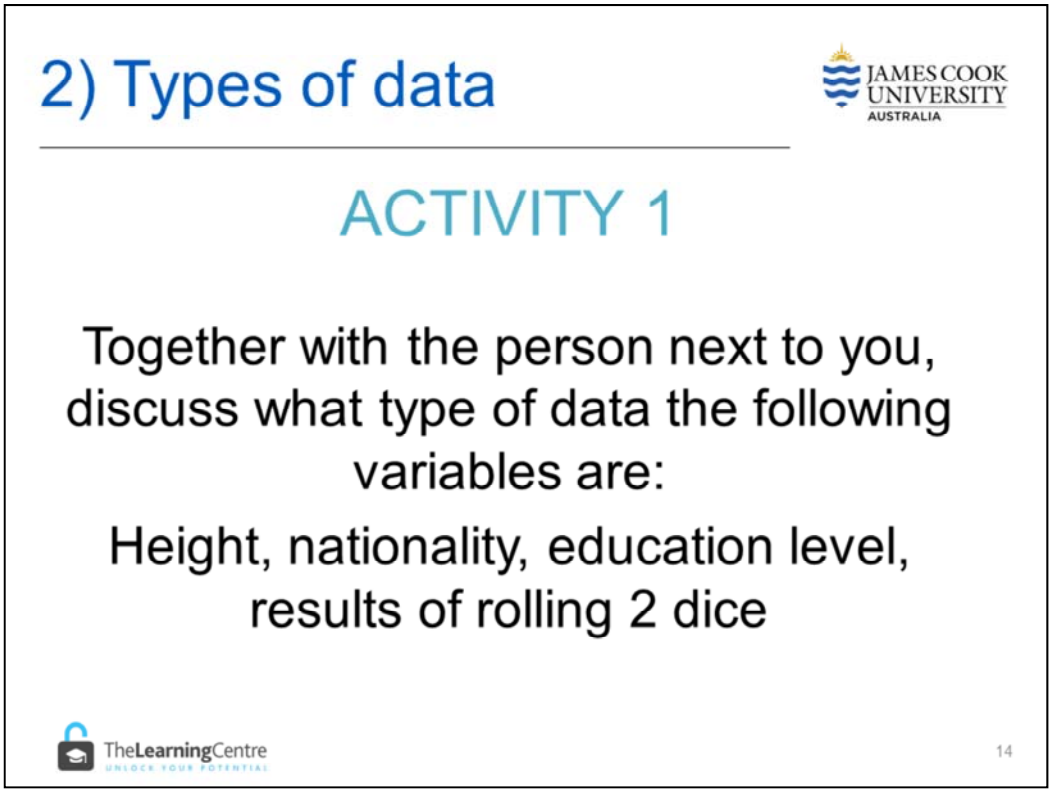
2) Types of data

JAMES COOK UNIVERSITY AUSTRALIA



TheLearningCentre
UNLOCK YOUR POTENTIAL

13

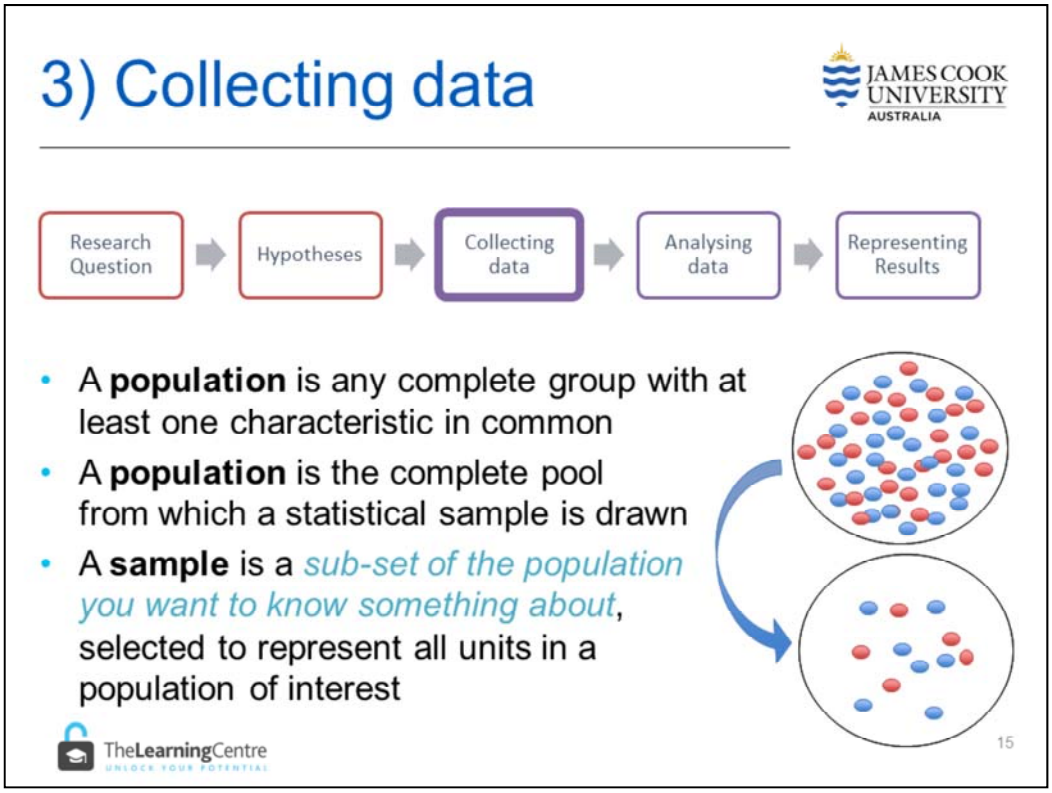


Height = continuous

Nationality = nominal

Education level = ordinal

results of rolling 2 dice = discrete



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.

Often it is not possible to measure/count every unit in a given population

Information from the sampled units is used to infer the characteristics for the entire population of interest

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

An *ideal* 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

TheLearningCentre
UNLOCK YOUR POTENTIAL

16

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

Random (probability) vs. non-probability sampling

Why random sampling? Why not?

- Random sampling reduces bias and sampling error

- Not possible; e.g. for ethical reasons
- Economic reasons (too expensive/accessibility of participants)

TheLearningCentre
UNLOCK YOUR POTENTIAL

17

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

Random (probability) vs. non-probability sampling

Why non-probability sampling? Why not?

- For ethical reasons (volunteer samples)
- For convenience (convenience sample)
- Increases chances of biases (self selection)

TheLearningCentre
UNLOCK YOUR POTENTIAL

18

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

- Simple Random Sampling
  - All units of the population are chosen at random and have the same chance of being selected

EXAMPLE
How heavy are dogs in Smithfield?
Assign each dog a number and use a random number table to select *random sample of dogs*

TheLearningCentre
UNLOCK YOUR POTENTIAL

3) Collecting data

- 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	
Population	All primary school students in Cairns
Groups	25 different primary schools in Cairns
Obtain Simple Random Sample	20 students from each of the primary schools
Sample	25 x 20 = 500 primary students selected

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

ACTIVITY 2

Together with the person next to you, think of a scenario in which **random sampling** would be possible and a scenario in which it would not.

TheLearningCentre
UNLOCK YOUR POTENTIAL

21

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

- Representative samples
  - the level at which the sample represents the population
  - If the sample does not represent the population, we are *limited in the types of conclusions that can be drawn*.
  - Sample should match the characteristics of the population
  - **Aim**: gain accuracy, but keep things small

TheLearningCentre
UNLOCK YOUR POTENTIAL

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

- Data collection methods
  - Questionnaires
  - Interviews
  - Observations
  - Physiological measurement
    - Self-report
    - Direct measurement

TheLearningCentre
UNLOCK YOUR POTENTIAL

3) Collecting data

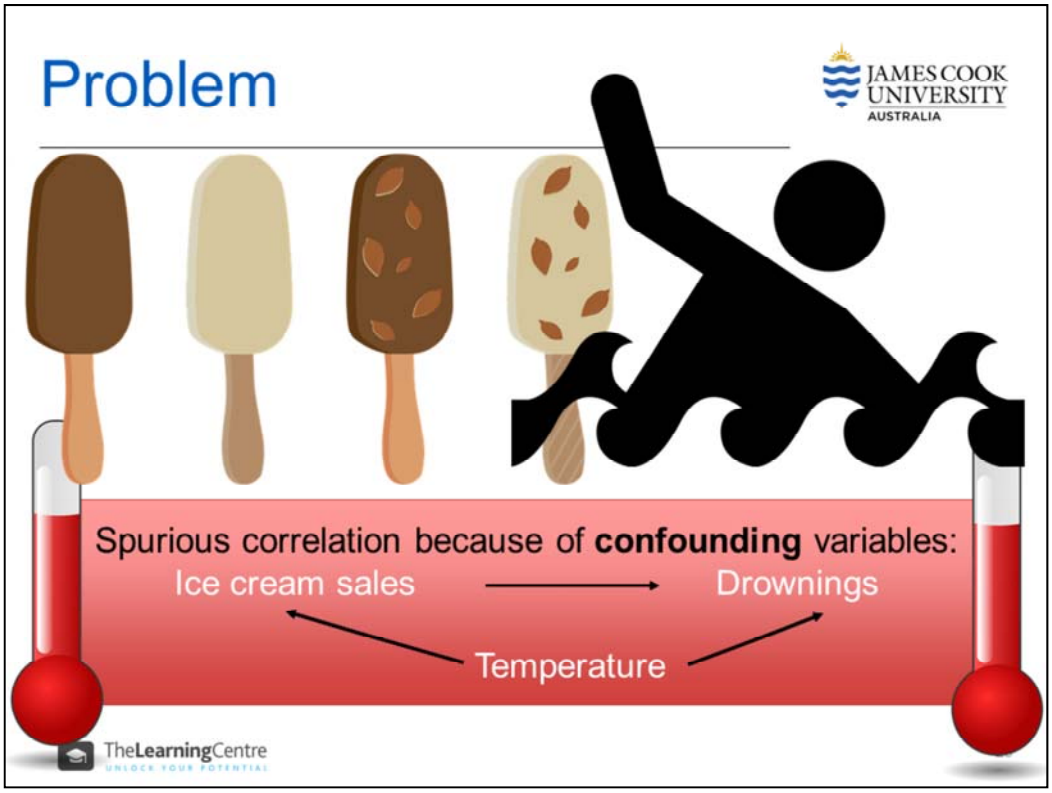
JAMES COOK UNIVERSITY AUSTRALIA

- Research methods & designs - The HOW and WHEN of collecting data

Research Designs / Research Methods	Longitudinal Design	Cross-Sectional Design
Experimental Research	Randomly assign subjects to groups. Measure **change**, **differences** and **differences in change** over time	Randomly assign subjects to groups. Measure **differences** at one point in time
Correlational Research	Determine the **relationship** between one variable over time *(i.e., 2 variables: time + one other)*	Determine if there is a **relationship** between the 2 variables at one point in time

TheLearningCentre
UNLOCK YOUR POTENTIAL

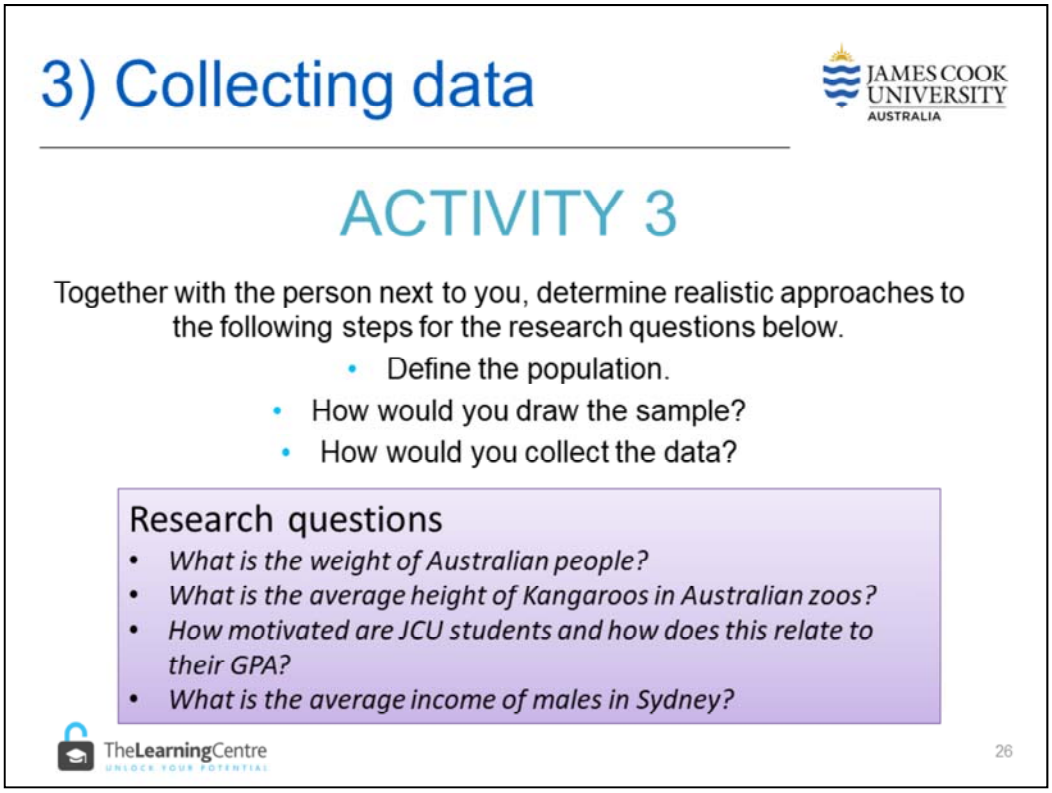
24



Research has shown that there is a strong positive correlation between ice cream sales and number of drownings.

Cause effect relationships cannot be determined conducting correlational research.

In experimental research, when possible confounding variables are controlled, cause effect relationships can be determined.



What is the weight of Australian people?

What is the average height of Kangaroos in Australian zoos?

How motivated are JCU students and how does this relate to their GPA?

What is the average income of males in Sydney?

3) Collecting data

JAMES COOK UNIVERSITY AUSTRALIA

Recap:

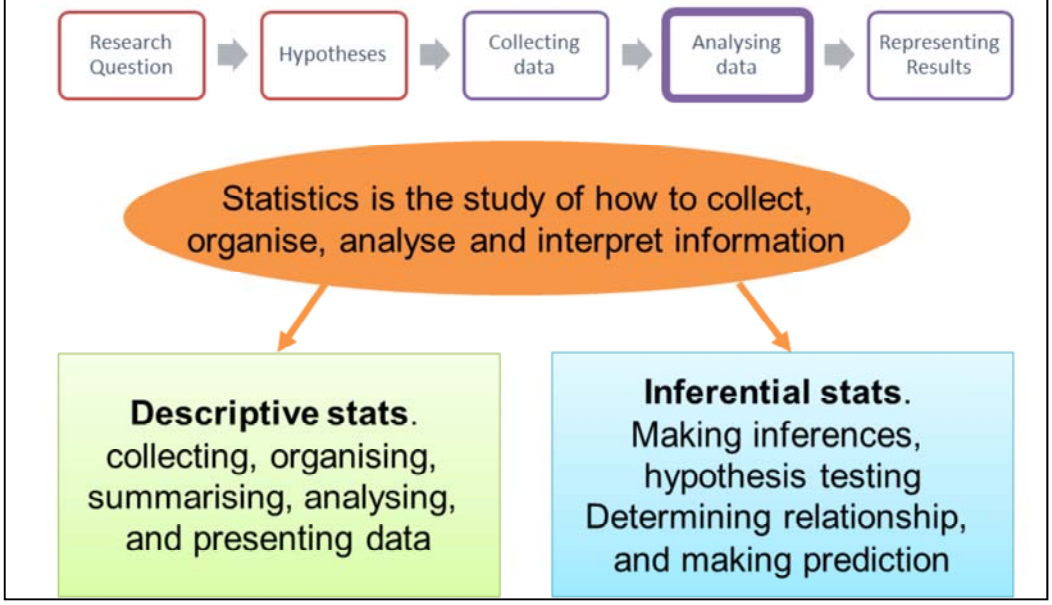
Statistics is the study of how to collect, organise, analyse and interpret information

- What do we need to consider during the data collection phase?
  - Research question
    - Types of data (variable types)
    - Population of interest
    - Sampling strategy & representativeness
    - Study design
      - Length of data collection (cross-sectional/longitudinal)
      - Comparison group? (e.g., experimental design)

TheLearningCentre UNLOCK YOUR POTENTIAL

27

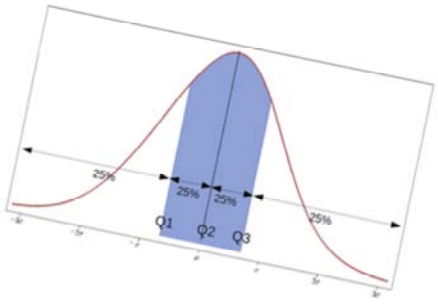
4) Analysing data



4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Describing basic features of the data
- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measures of spread
  - Range
  - Quartiles
  - Variance and standard deviation



TheLearningCentre
UNLOCK YOUR POTENTIAL

29

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Mean: Arithmetic average

$$Mean = \frac{Sum\ of\ all\ data\ values}{Number\ of\ data\ values}$$

$$\bar{x} = \frac{\sum x_i}{n}$$ where

$\bar{x}$	(read as x bar) is the mean of the set of x values
$\sum x_i$	is the sum of all x values
n	is the number of x values

TheLearningCentre UNLOCK YOUR POTENTIAL

30

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Median
  - Middle value of rank ordered data
  - Value that separates the higher half of a data set from the lower half
  - Can be found by arranging all values from lowest to highest and determining the value in the middle
  - 7,9,2,1,3,7,8,5,1
  - 1,1,2,3,5,7,7,8,9

1,1,2,3,7,7,8,9

$$\frac{3+7}{2} = 5$$

TheLearningCentre
UNLOCK YOUR POTENTIAL

31

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Mode
  - The most frequent measurement
  - 12,11,15,12,12,11,14,17,15,12,13

Participants' age	Frequency
11	2
12	4
13	1
14	1
15	2
17	1

TheLearningCentre UNLOCK YOUR POTENTIAL

32

- If no number is repeated in the data set, there is no mode
- Bimodal when two most frequent measurements

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

Measures of spread/ dispersion

- *Describes variability* in a sample
- Used in conjunction with a measure of central tendency to provide overall description of data

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Range
  - Simplest measure of spread
  - Difference between the largest value and the smallest value of a dataset
  - Range = maximum value – minimum value

23 46 33 **12** 44 31 29 15 **47** 37 22 34 35 41 36

Min Value

Max Value

TheLearningCentre
UNLOCK YOUR POTENTIAL

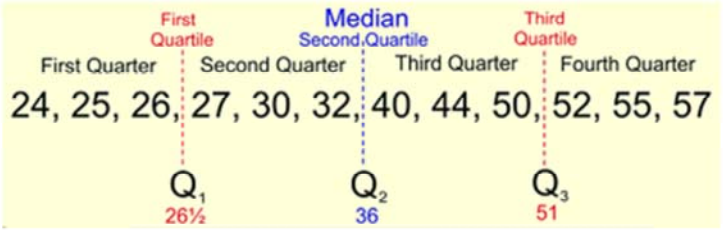
Range = 47 - 12 = 35

34

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Quartiles
  - Divide data into four groups - each with an equal number of data points
  - Ranked data – arranged into ascending order of magnitude



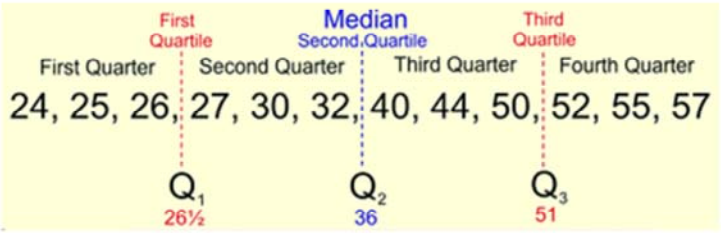
TheLearningCentre UNLOCK YOUR POTENTIAL

Image accessed: http://ecstep.com/types-of-data/ 35

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Interquartile Range
  - Difference between the third quartile, Q3 and the first quartile, Q1
  - IQR = Q3 - Q1
  - Range for the middle 50 % of data



IQR = 51 – 26.5 = 24.5

TheLearningCentre UNLOCK YOUR POTENTIAL

Image accessed: http://ecstep.com/types-of-data/ 36

4) Analysing data: Part I – Descriptive Statistics

- Standard deviation
  - Standard deviation is a standardised numerical value which indicates how 'spread out' a group of data points are
  - Standards deviation is derived from the difference between the value of each observation and the mean
  - If individual observations vary greatly from the group mean, the variance is big; and vice versa

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where: s = the standard deviation of the sample
x_i = the measurement of each data unit in the sample
$\bar{x}$ = the sample mean
n = the size of the sample (the number of data values)

TheLearningCentre
UNLOCK YOUR POTENTIAL

38

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i
19
22
26
18
19
20
28
24

TheLearningCentre
UNLOCK YOUR POTENTIAL

39

What do we have to calculate first?

The mean

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i
19
22
26
18
19
20
28
24

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{19 + 22 + 26 + 18 + 19 + 20 + 28 + 24}{8}$$

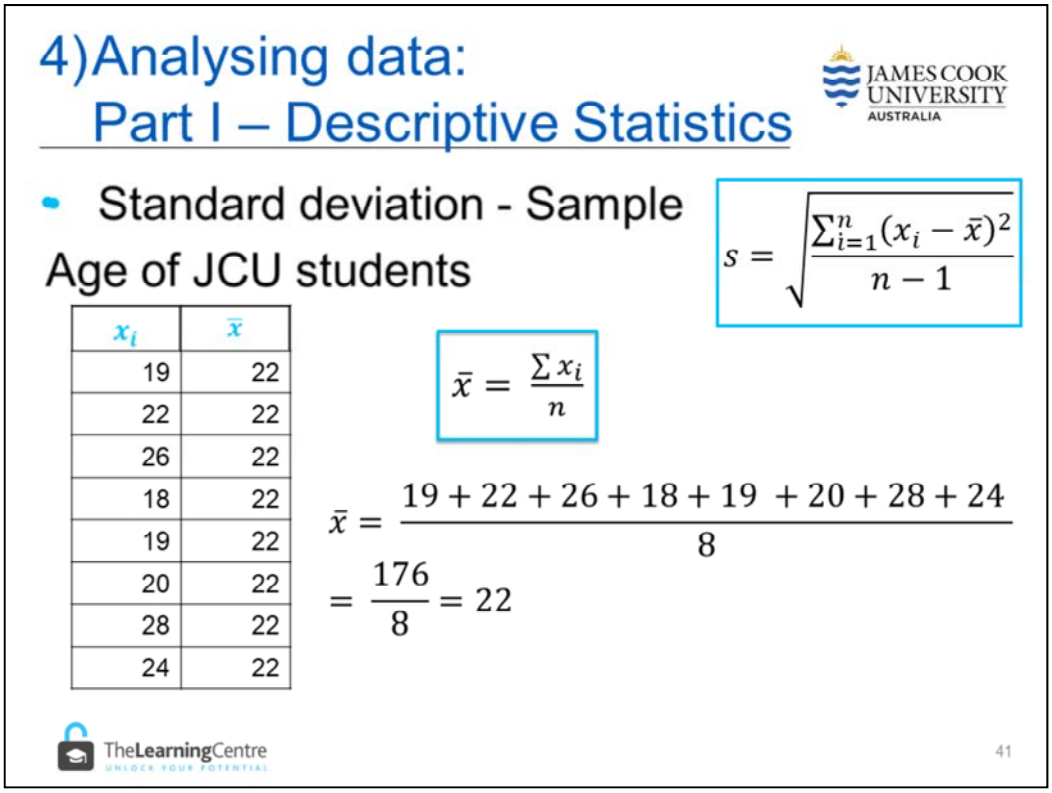
$$= \frac{176}{8} = 22$$

TheLearningCentre UNLOCK YOUR POTENTIAL

40

What do we have to calculate first?

The mean



What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$
19	22	
22	22	
26	22	
18	22	
19	22	
20	22	
28	22	
24	22	

TheLearningCentre UNLOCK YOUR POTENTIAL

42

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$
19	22	-3
22	22	0
26	22	4
18	22	-4
19	22	-3
20	22	-2
28	22	6
24	22	2

TheLearningCentre UNLOCK YOUR POTENTIAL

43

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

What is the third step?

Square each value for x_i - $(x)^-$

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
19	22	-3	
22	22	0	
26	22	4	
18	22	-4	
19	22	-3	
20	22	-2	
28	22	6	
24	22	2	

TheLearningCentre UNLOCK YOUR POTENTIAL

44

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

What is the third step?

Square each value for x_i - $(x)^-$

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
19	22	-3	9
22	22	0	0
26	22	4	16
18	22	-4	16
19	22	-3	9
20	22	-2	4
28	22	6	36
24	22	2	4

TheLearningCentre UNLOCK YOUR POTENTIAL

45

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

What is the third step?

Square each value for x_i - $(x)^{-}$

What is the fourth step?

Add up all squared values

4) Analysing data: Part I – Descriptive Statistics

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
19	22	-3	9
22	22	0	0
26	22	4	16
18	22	-4	16
19	22	-3	9
20	22	-2	4
28	22	6	36
24	22	2	4
			94

TheLearningCentre

46

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

What is the third step?

Square each value for x_i - $(x)^{-}$

What is the fourth step?

Add up all squared values

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Standard deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Age of JCU students

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
19	22	-3	9
22	22	0	0
26	22	4	16
18	22	-4	16
19	22	-3	9
20	22	-2	4
28	22	6	36
24	22	2	4
			94

$$s = \sqrt{\frac{94}{7}} = 3.66$$

TheLearningCentre UNLOCK YOUR POTENTIAL

47

What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

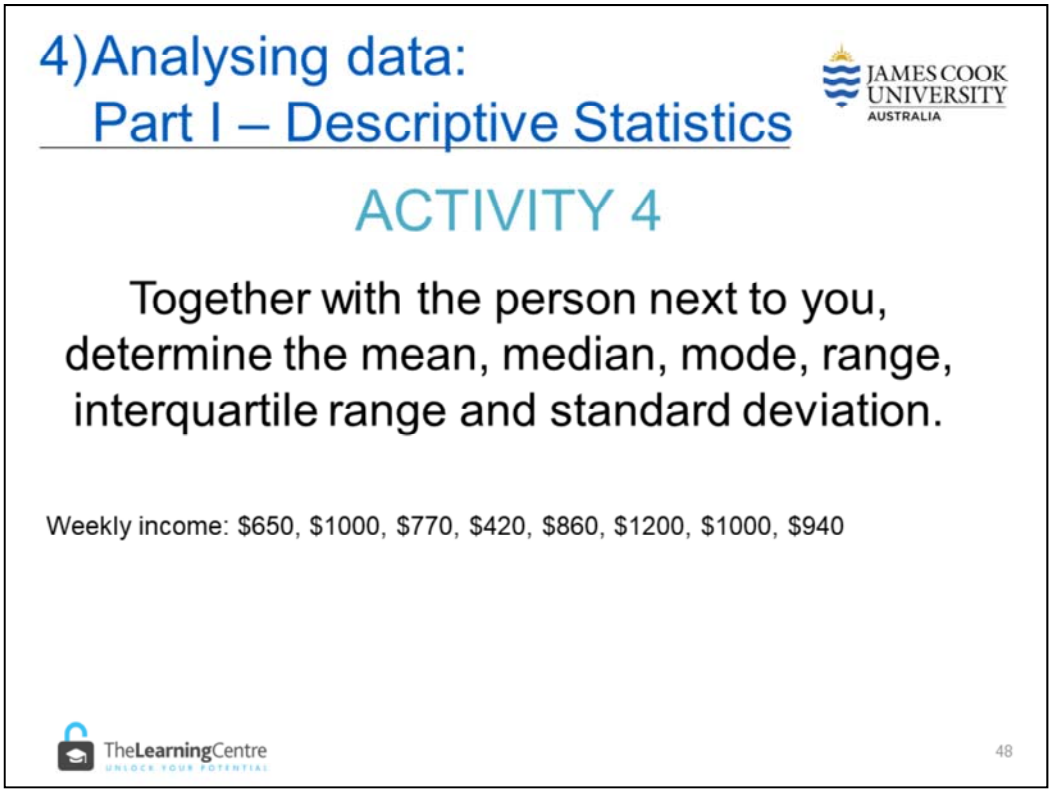
What is the third step?

Square each value for x_i - $(x)^{-}$

What is the fourth step?

Add up all squared values

The we can put all values in the formula and calculate the standard deviation.



Mean: $855

Mode: $1000

Median: 420, 650, 770, 860, 940, 1000, 1000, 1200 → (860 + 940)/2 = 900
Q1 = (650 + 770)/2 = 710 Q3 = 1000

Range = maximum value – minimum value: 1200 - 420 = 780

Interquartile range = Q3 - Q1 = 1000 - 710 = 290

Standard deviation: 241.07

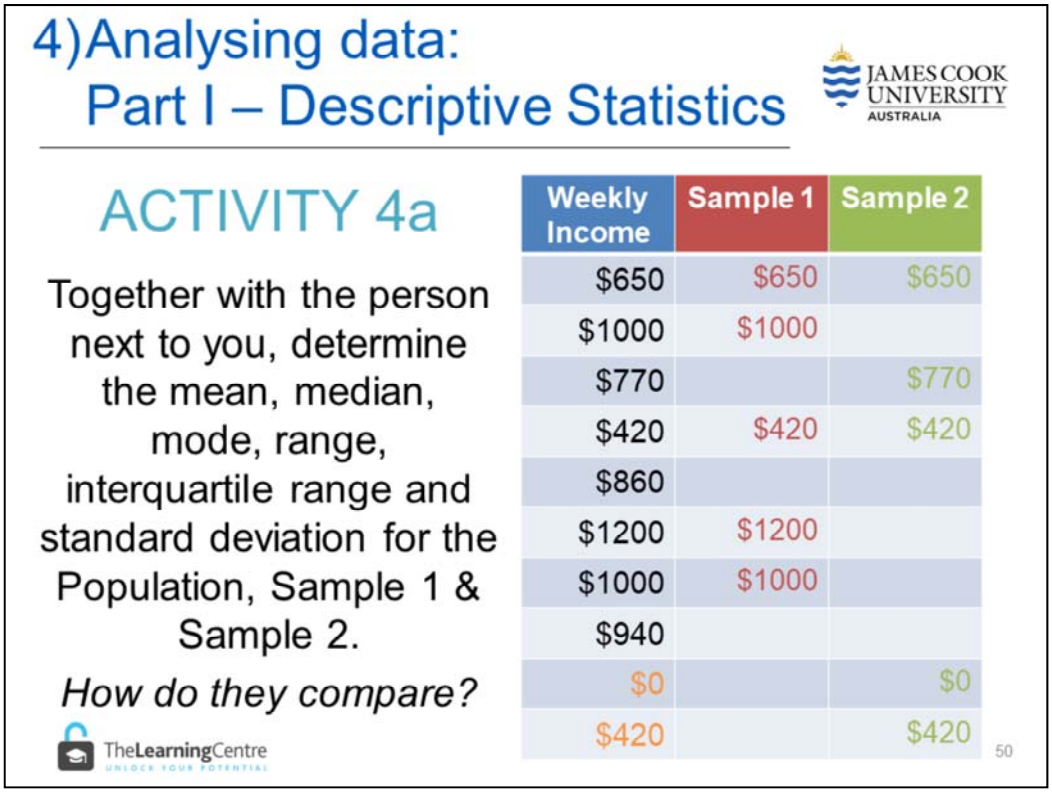
4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

ACTIVITY 4 - Answers

Descriptive Statistic	Value
Mean	$855
Median	$900
Mode	$1000
Range	$780
IQR	$290
SD	241.07

TheLearningCentre
UNLOCK YOUR POTENTIAL



ORIGINAL POPULATION:

Mean: $855

Mode: $1000

Median: 420, 650, 770, 860, 940, 1000, 1000, 1200 → (860 + 940)/2 = 900
Q1 = (650 + 770)/2 = 710 Q3 = 1000

Range = maximum value – minimum value: 1200 - 420 = 780

Interquartile range = Q3 - Q1 = 1000 - 710 = 290

Standard deviation: 241.07

4) Analysing data: Part I – Descriptive Statistics

JAMES COOK UNIVERSITY AUSTRALIA

ACTIVITY 4a - Answers

Descriptive Statistic	Original population (n = 8)	New population (n = 10)	Sample 1 (n = 5)	Sample 2 (n = 5)
Mean	$855	$726	$854	$452
Median	$900	$815	$1000	$420
Mode	$1000	$420 & $1000	$1000	$420
Range	$780	$1200	$780	$770
IQR	$290	$580	$565	$500
SD	241.07	359.11	313.18	294.40

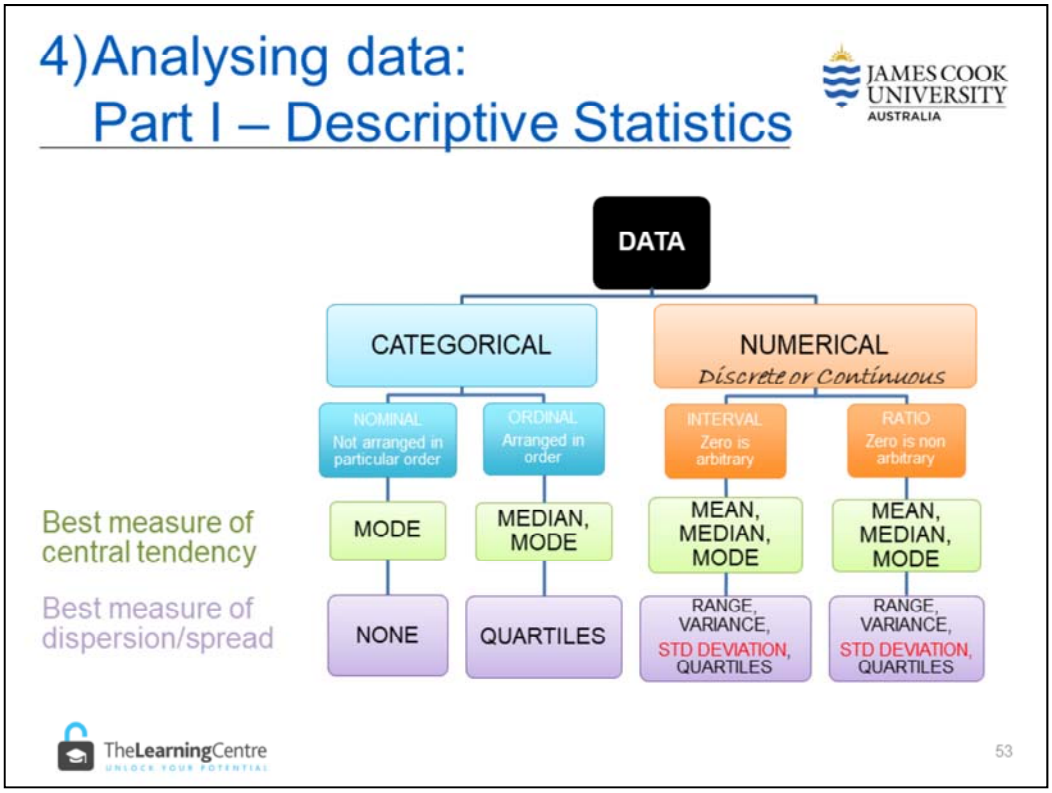
→ **Each sample within a population can result in different values for descriptive statistics.**

TheLearningCentre
UNLOCK YOUR POTENTIAL

51

Sections 1-4: Recap/Summary

- The goal of inferential statistics is to take observable data from a sample of a larger population and make inferences about the population.
- Samples can tell us something about the population to the extent that the sampling is representative of the population.



What do we have to calculate first?

The mean

What is the second step?

Subtract the mean of the sample from the measurement of each data unit in the sample

What is the third step?

Square each value for x_i - $(x)^{-}$

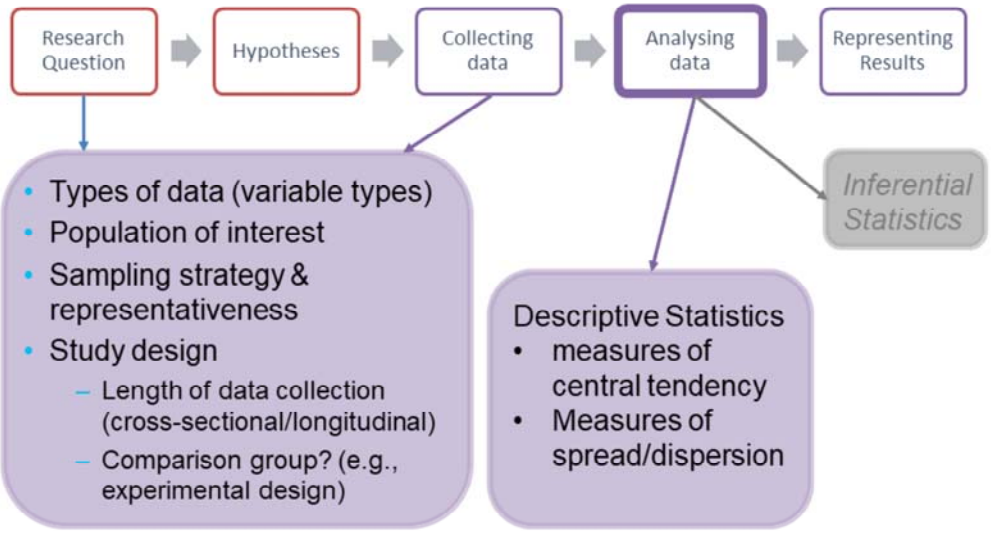
What is the fourth step?

Add up all squared values

The we can put all values in the formula and calculate the standard deviation.

Sections 1-4 Recap/Summary

JAMES COOK UNIVERSITY AUSTRALIA



TheLearningCentre
UNLOCK YOUR POTENTIAL

54


JAMES COOK UNIVERSITY
AUSTRALIA
Lunch
TheLearningCentre
UNLOCK YOUR POTENTIAL
Cairns
Singapore
Townsville

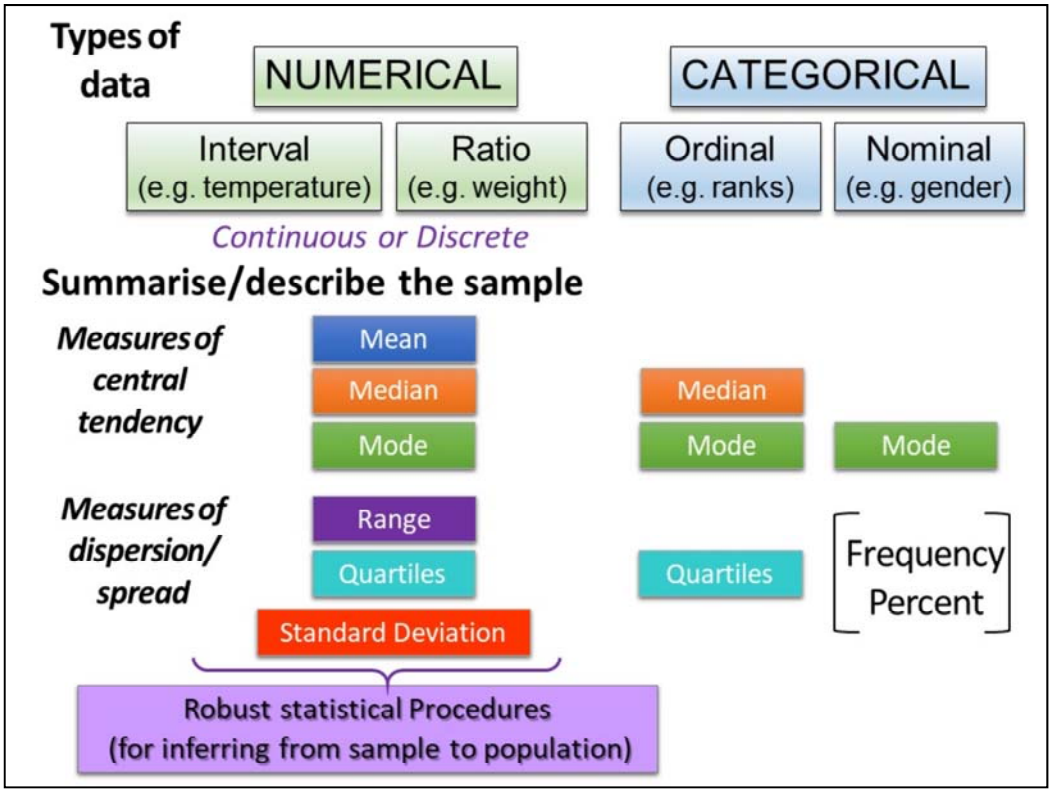
Afternoon Session

JAMES COOK UNIVERSITY AUSTRALIA

Inferential Statistics

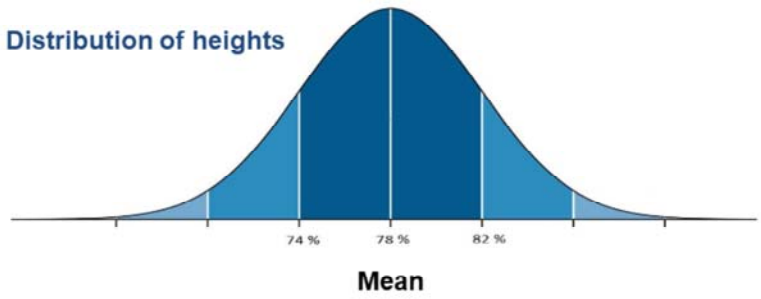
1) Introduction: Statistics and the research process
2) Types of data
3) Collecting data
4) Analysing data: Part I – Descriptive Statistics
5) Data characteristics: Normal distribution
6) Analysing data: Part II – Inferential Statistics
7) Representing results

TheLearningCentre
UNLOCK YOUR POTENTIAL


Types of data
NUMERICAL
CATEGORICAL
Interval (e.g. temperature)
Ratio (e.g. weight)
Ordinal (e.g. ranks)
Nominal (e.g. gender)
Continuous or Discrete
Summarise/describe the sample
Measures of central tendency
Mean
Median
Mode
Median
Mode
Mode
Measures of dispersion/ spread
Range
Quartiles
Standard Deviation
Quartiles
Frequency
Percent
Robust statistical Procedures
(for inferring from sample to population)

5) Data characteristics: Normal distribution

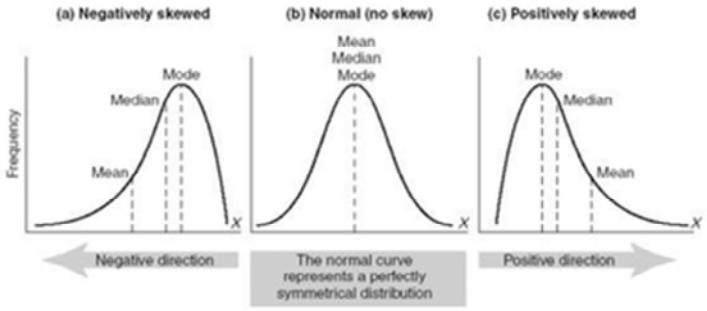
- Definition:
  - The distribution is bell shaped and symmetrical around the mean
  - The total area under the curve is equal to 1
- Importance: allows for Parametric tests



5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- Determining normal distribution
  1. Central Limit Theorem
  2. Descriptive statistics
     a) **The mean, mode and median are the same value**
     b) Standard deviation

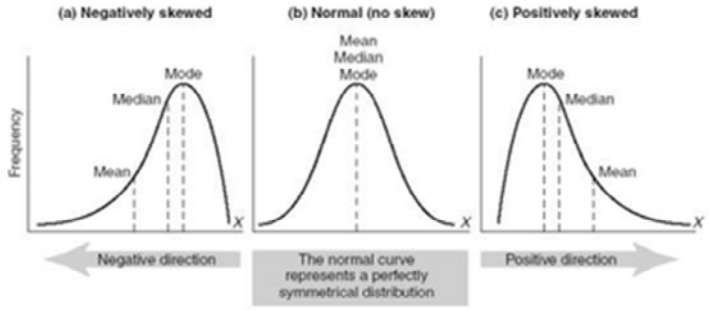


Examples of normal and skewed distributions

TheLearningCentre UNLOCK YOUR POTENTIAL

5) Data characteristics: Normal distribution

Descriptive Statistic	Original population (n = 8)	New population (n = 10)	Sample 1 (n = 5)	Sample 2 (n = 5)
Mean	$855	$726	$854	$452
Median	$900	$815	$1000	$420
Mode	$1000	$420 & $1000	$1000	$420



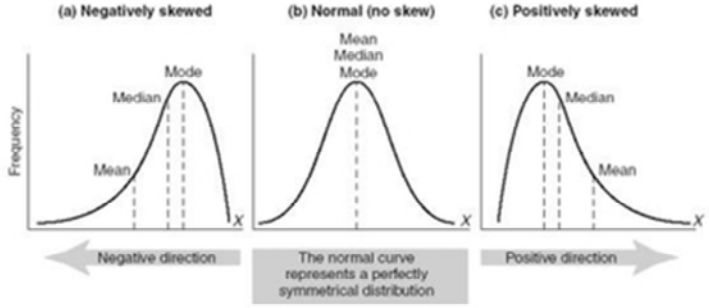
Examples of normal and skewed distributions

TheLearningCentre
UNLOCK YOUR POTENTIAL

60

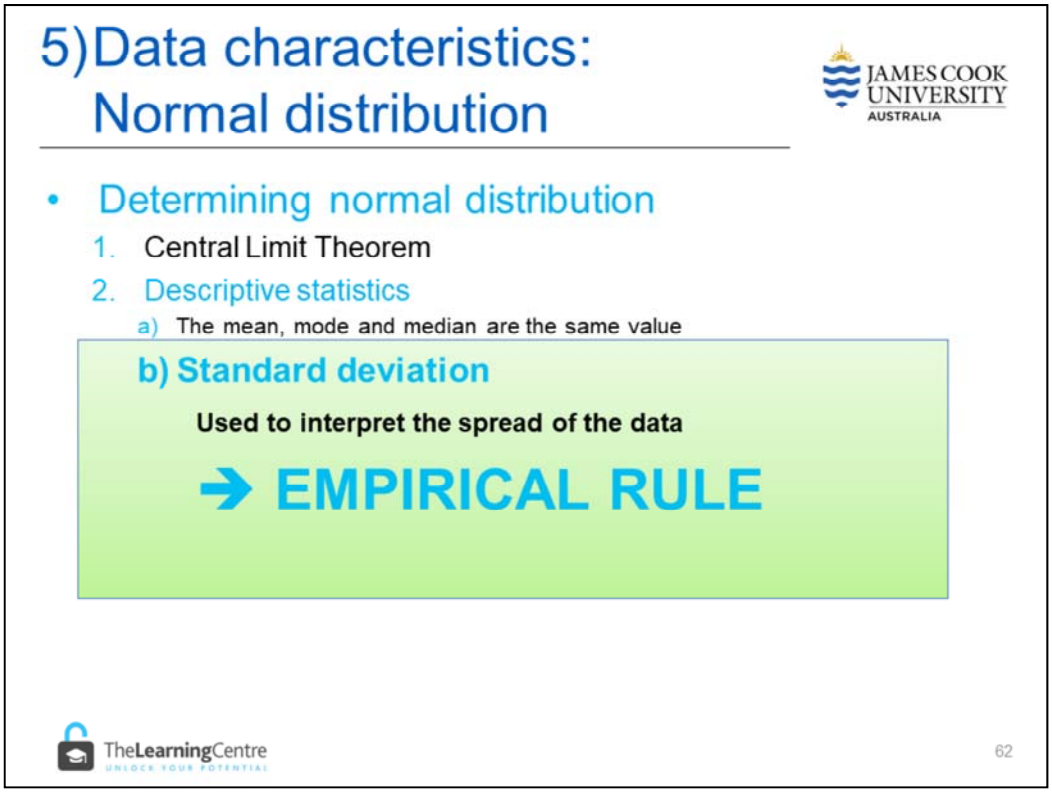
5) Data characteristics: Normal distribution

- Determining normal distribution
  1. Central Limit Theorem
  2. Descriptive statistics
     a) The mean, mode and median are the same value
     b) **Standard deviation**



Examples of normal and skewed distributions

TheLearningCentre
UNLOCK YOUR POTENTIAL

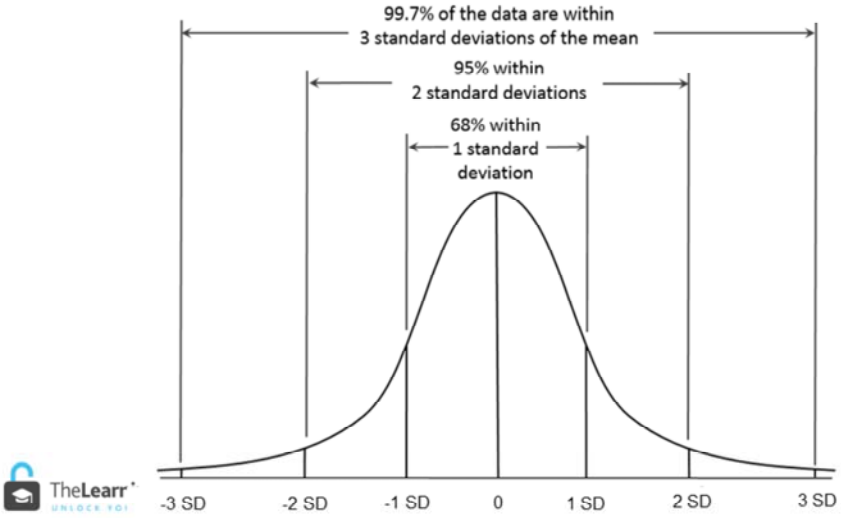


What can the standard deviation be used for?

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- EMPIRICAL RULE



TheLearn

63

The empirical rule tells us approximately 68 % of data values will fall within 1 standard deviation of the mean

95 % of data values will fall within 2 standard deviations of the mean

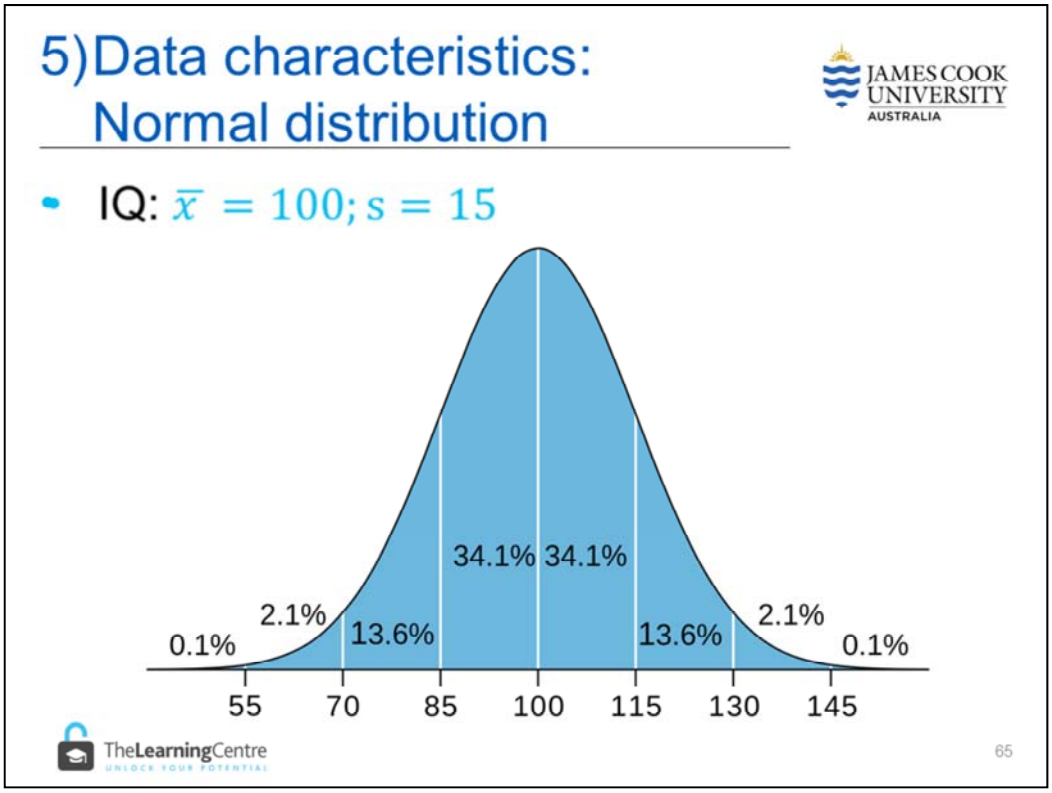
99.7 % of data values will fall within 3 standard deviations of the mean

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- What can you do with a normal curve?
  - Estimate the probability of a certain data point appearing in the dataset

TheLearningCentre
UNLOCK YOUR POTENTIAL



What does this mean?

68 % of the population has an IQ between 85 and 115.

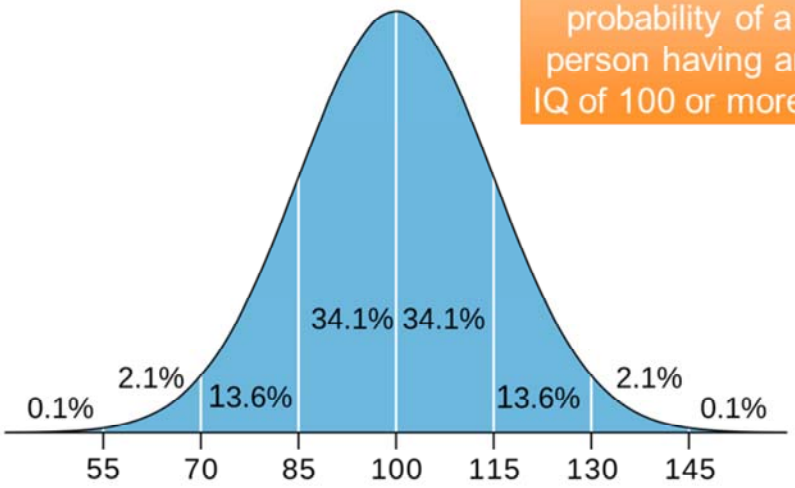
Only a very small percentage are highly gifted (130 and over) only around 2 %.`

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$

What is the probability of a person having an IQ of 100 or more?



TheLearningCentre
UNLOCK YOUR POTENTIAL

66

0.5 (50 %)

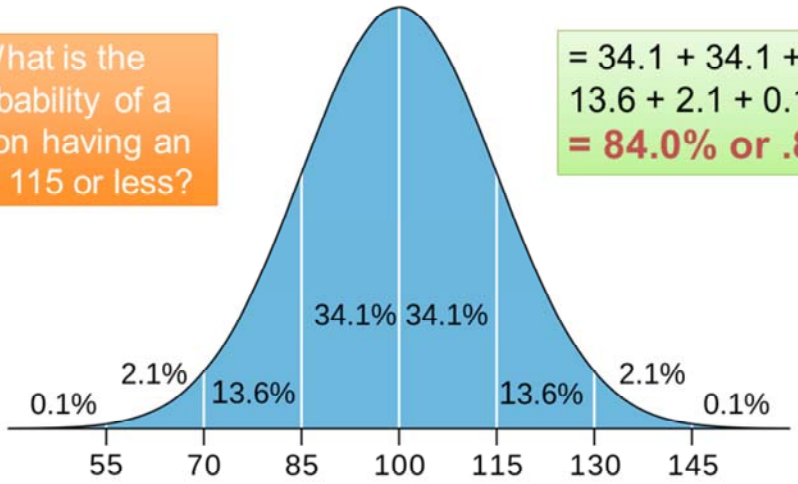
5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$

What is the probability of a person having an IQ of 115 or less?

= 34.1 + 34.1 + 13.6 + 2.1 + 0.1
= 84.0% or .81



TheLearningCentre
UNLOCK YOUR POTENTIAL

67

0.84 (84 %)

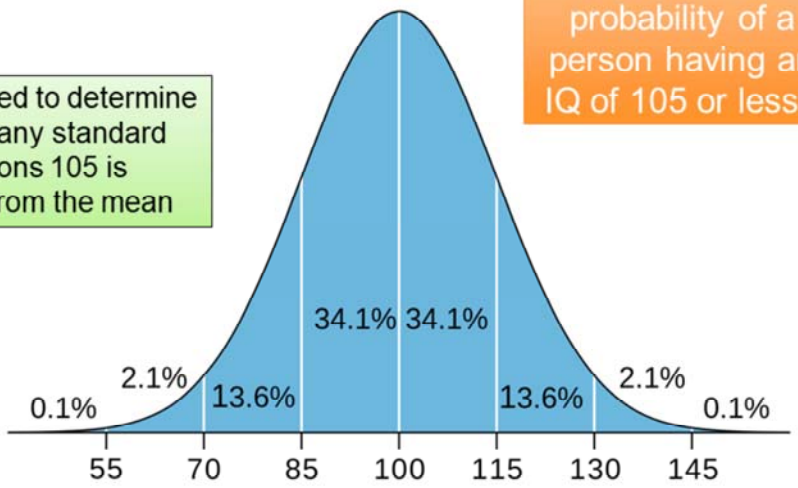
5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$

What is the probability of a person having an IQ of 105 or less?

We need to determine how many standard deviations 105 is away from the mean



TheLearningCentre UNLOCK YOUR POTENTIAL

68

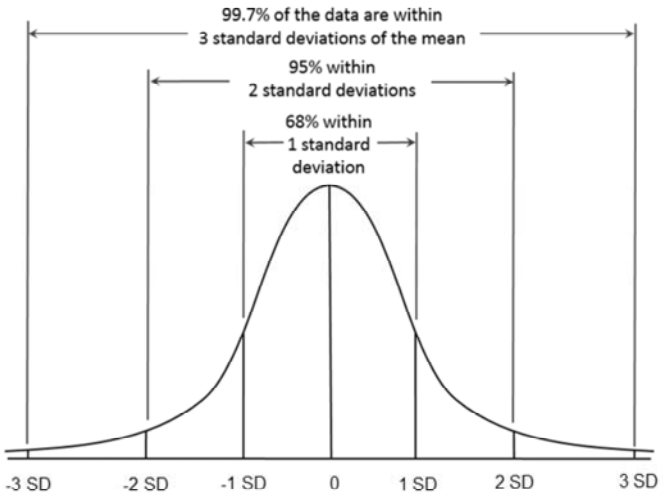
5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- Standard normal distribution

$\bar{x} = 0$

$s = 1$



TheLearningCentre
UNLOCK YOUR POTENTIAL

69

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$
- The z-score

What is the probability of a person having an IQ of 105 or less?

$$z = \frac{x - \bar{x}}{s}$$

x = the normally distributed variable of interest
$\bar{x}$ = the mean of the normal distribution
s = the standard deviation of the normal distribution
z = the z-score (the number of standard deviations between x and $\bar{x}$)

TheLearningCentre
UNLOCK YOUR POTENTIAL

70

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$
- The z-score

What is the probability of a person having an IQ of 105 or less?

$$z = \frac{x - \bar{x}}{s} = \frac{105 - 100}{15} = \frac{5}{15} = 0.33$$

TheLearningCentre
UNLOCK YOUR POTENTIAL

71

5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

- IQ: $\bar{x} = 100; s = 15$
- The z-score

What is the probability of a person having an IQ of 105 or less?

$$z = \frac{x - \bar{x}}{s} = \frac{105 - 100}{15} = \frac{5}{15} = 0.33$$

- Now that you have the standard z-score (0.33), use a z-score table to determine the probability

TheLearningCentre
UNLOCK YOUR POTENTIAL

72

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Image accessed: https://statistics.laerd.com

The probability of having an IQ of 105 or less is 0.6255 (62.55%)

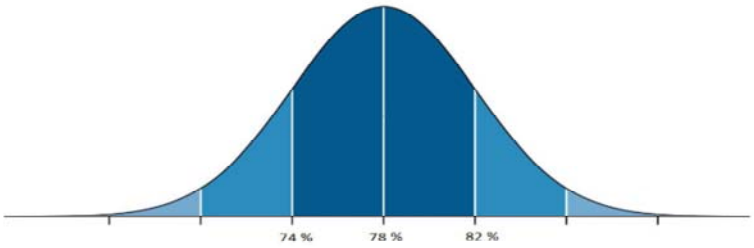
5) Data characteristics: Normal distribution

JAMES COOK UNIVERSITY AUSTRALIA

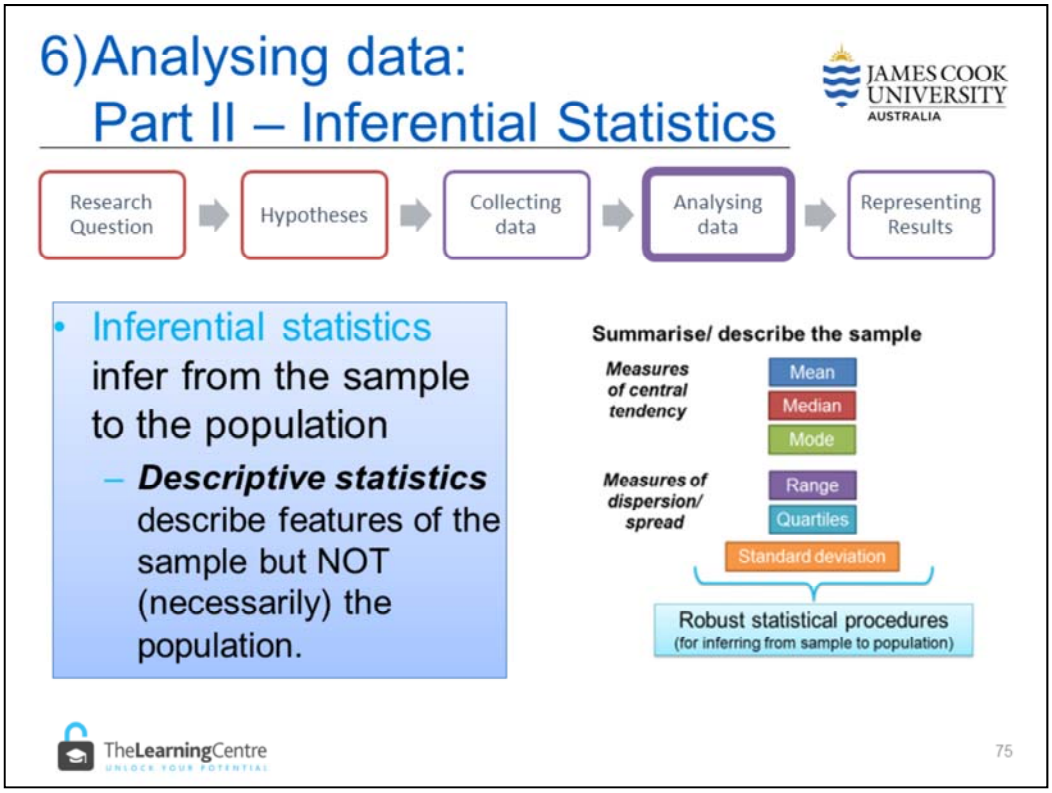
SUMMARY

- Having a normal distribution for a numerical variable allows for robust parametric statistical procedures
  - Enables us to calculate the probability a particular value of the numerical variable will occur in the population.

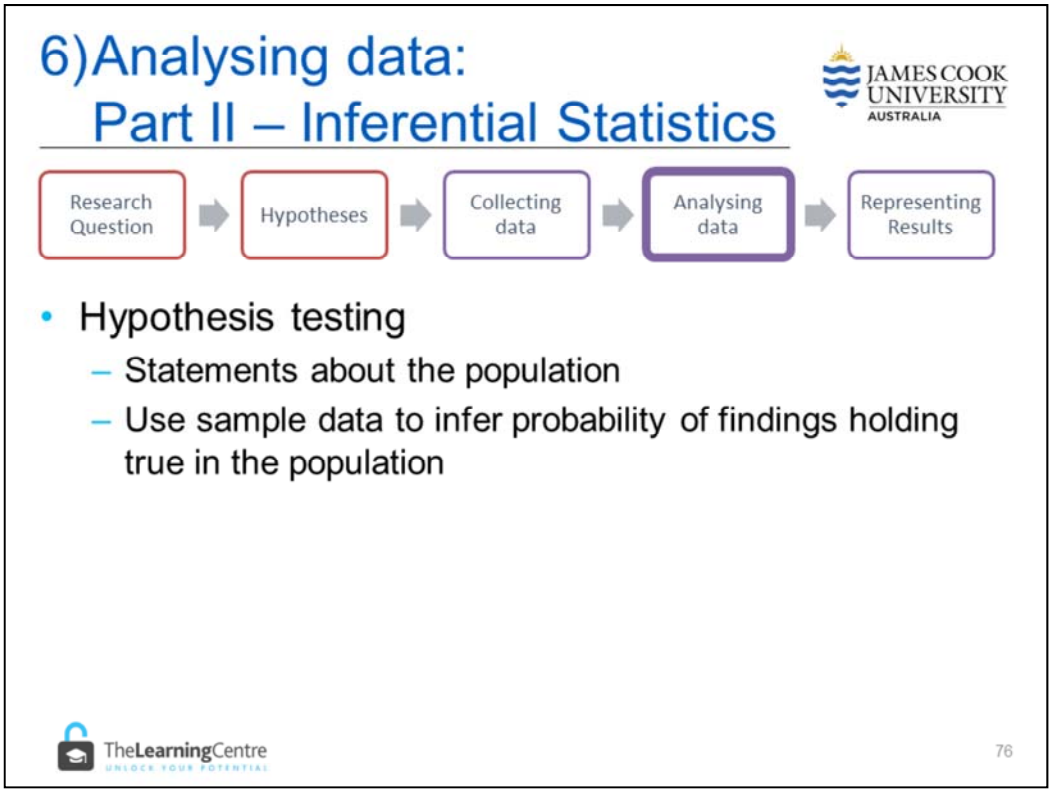
→ Inferential statistics



TheLearningCentre
UNLOCK YOUR POTENTIAL



Hypothesis testing aims to make a statistical conclusion about accepting or not accepting the hypothesis



Hypothesis testing aims to make a statistical conclusion about accepting or not accepting the hypothesis

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Broad steps for hypothesis testing
  - **Determine appropriate test**
  - Perform test
  - Compare results to an established significance level (usually $p<0.05$)
  - Interpret results & draw conclusions.

TheLearningCentre
UNLOCK YOUR POTENTIAL

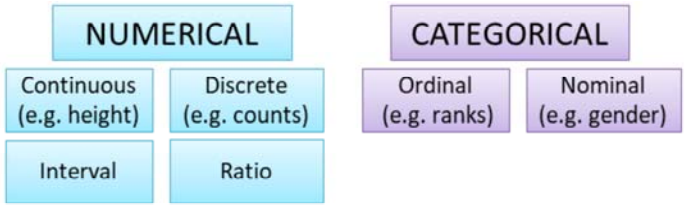
77

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Broad steps for hypothesis testing
  - **Determine appropriate test**
    1. Identify variable types

Types of data



    2. Identify what is being tested
    3. Identify the hypothesis type
    4. Form hypotheses

TheLearningCentre UNLOCK YOUR POTENTIAL

78

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Examples

Two studies	Exam prep time (min) & Exam score (%)	Plant size (cm) & Growing condition (fertilizer or not)
1. Define variables		
2. What is tested?		
3. Hypothesis type		
4. Possible hypotheses		

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

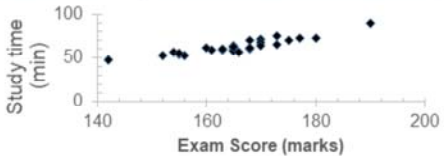
- Examples

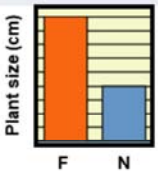
Two studies	Exam prep time (min) & Exam score (%)	Plant size (cm) & Growing condition (fertilizer or not)
1. Define variables	2 numerical	1 numerical & 1 categorical
2. What is tested?		
3. Hypothesis type		
4. Possible hypotheses		

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

Two studies	Exam prep time (min) & Exam score (marks)	Plant size (cm) & Growing condition (fertilizer or not)
1. Define variables	2 numerical	1 numerical & 1 categorical
2. What is tested?	Relationship between two variables	Difference between two groups
3. Hypothesis type	Correlational	Difference
4. Possible hypotheses		





TheLearningCentre
UNLOCK YOUR POTENTIAL

81

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Examples

Two studies	Exam prep time (min) & Exam score (%)	Plant size (cm) & Growing condition (fertilizer or not)
1. Define variables	2 numerical	1 numerical & 1 categorical
2. What is tested?	Relationship between two variables	Difference between two groups
3. Hypothesis type	Correlational	Difference
4. Possible (conceptual) hypotheses	The more time spent studying for the exam, the higher the exam score.	Fertilized plants are larger than unfertilized plants.

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Broad steps for hypothesis testing
  - **Determine appropriate test**

Type of Problem	Type of Data	Parametric Methods	Nonparametric Methods
Comparison of groups	One group (compared to a reference value)	z-test, t-test	Chi-squared test, Kolmogorov-Smirnov test
	Two independent groups	t-test, z-test, analysis of variance	Wilcoxon's signed rank test, median test, chi-squared test, Kolmogorov-Smirnov test, Mann-Whitney test
	Two paired or related groups	Paired t-test, z-test	Wilcoxon rank sum test, sign test
	Three or more groups	Analysis of variance, z-test	Kruskall-Wallis test, Friedman two-way analysis of variance by ranks
Association	One sample	Least-squares correlation analysis	Spearman rank correlation coefficient, Kendall's rank correlation coefficient (tau)
	More than one sample[b]	Regression analysis or logistical regression	Chi-squared test of independence

TheLearningCentre UNLOCK YOUR POTENTIAL

Source: Kuzon, W.M., Urbanchek, M.G., & McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery* 37(3), p.267.

83

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

Types of hypotheses –

- Conceptual
- Operational
- Statistical
  - Null hypothesis - H_0
    - States there is no statistical significance between two variables in the hypothesis
    - Believed to be true unless there is overwhelming evidence to the contrary
  - Alternative hypothesis - H_1 or H_a
    - States that there is a statistical significance between two variables
    - Holds true if the null hypothesis is rejected

TheLearningCentre
UNLOCK YOUR POTENTIAL

84

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

TheLe

85

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Types of hypotheses

	Correlation Hypothesis	Difference Hypothesis
Non-directional	There is no statistically significant relationship between time spent studying and exam outcomes H_0: $r = 0$ There is a statistically significant relationship between time spent studying and exam outcomes H_1: $r \neq 0$	Fertilised plants and non-fertilised plants do not significantly differ in size H_0: $\mu_1 = \mu_2$ Fertilised plants and non-fertilised plants significantly differ in size H_1: $\mu_1 \neq \mu_2$
Directional	There is no statistically significant positive relationship between time spent studying and exam outcomes H_0: $r \leq 0$ There is a statistically significant positive relationship between time spent studying and exam outcomes H_1: $r > 0$	Fertilised plants are not significantly bigger than non-fertilised plants H_0: $\mu_1 \leq \mu_2$ Fertilised plants are significantly bigger than non-fertilised plants H_1: $\mu_1 > \mu_2$

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

ACTIVITY 5

Together with the person next to you, outline the research process for the following research question:

ARE VIDEO GAMES BENEFICIAL FOR A CHILD'S COGNITIVE DEVELOPMENT?

Consider: Types of variables, sample, research design & method, possible conceptual hypothesis & corresponding statistical hypotheses

TheLearningCentre
UNLOCK YOUR POTENTIAL

87

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Broad steps for hypothesis testing
  - Determine appropriate test
  - Perform test
  - Compare results to an established significance level (usually $p<0.05$)
  - Interpret results & draw conclusions.

TheLearningCentre
UNLOCK YOUR POTENTIAL

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Statistical significance
  - Significance is indicated by a p-value less than .05 ($p < .05$)
  - Significant means that the outcome was unlikely to have occurred by chance
  - If our result is statistically significant, we have enough evidence to reject the null hypothesis

Hypothesis testing example

JAMES COOK UNIVERSITY AUSTRALIA

Stating the statistical hypotheses

- I have an assignment due for my subject. My hypothesis is that it will take an average student 6 days to complete the assignment. I want to test this hypothesis – that the population mean, μ, is equal to six days.
  - I.e.,

$$H_0 : \mu = 6.0 \text{ days}$$

$$H_1 : \mu \neq 6.0 \text{ days}$$

TheLearningCentre UNLOCK YOUR POTENTIAL

Image accessed: http://joindrpam.com/write-assignment/

90

Hypothesis testing Example

JAMES COOK UNIVERSITY AUSTRALIA

Conducting the test:

- To conduct the test, I gather a sample of people who have completed the assignment in the past and calculate the average number of days it took them to complete it. Suppose the sample mean is 6.1 days. The hypothesis test will tell me whether 6.1 days is significantly different from 6.0 days.

TheLearningCentre
UNLOCK YOUR POTENTIAL

Image accessed: http://joindrpam.com/write-assignment/

91

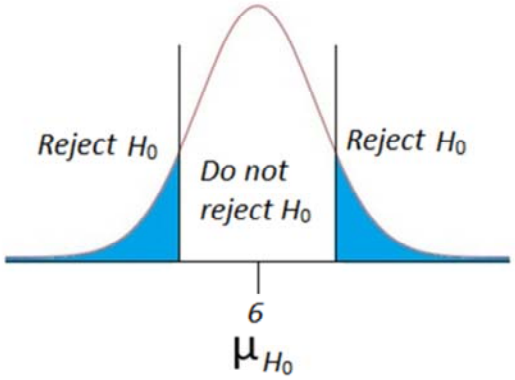
Two –Tail Hypothesis Test

JAMES COOK UNIVERSITY AUSTRALIA

- The curve represents the sampling distribution of the mean for the number of days it takes to complete the assignment

$H_0 : \mu = 6.0$ days

$H_1 : \mu \neq 6.0$ days



Mean number of days to complete assignment

TheLearningCentre UNLOCK YOUR POTENTIAL

92

The mean of the population – assumed to be 6 days according to the null hypothesis, is the mean of the sampling distribution

Two –Tail Hypothesis Test

JAMES COOK UNIVERSITY AUSTRALIA

There are only two statements we can make about the null hypothesis:

- Reject the null hypothesis
- Do not reject the null hypothesis
  - Because we use sample data to infer to the population, we cannot accept the null, only fail to reject it.

To remember, use the analogy of the legal system. When a jury or judge finds a defendant "not guilty," they are not saying the defendant is innocent. They are saying there is not enough evidence to prove guilt.

TheLearningCentre UNLOCK YOUR POTENTIAL

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

93

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Interpreting results – draw conclusions
  - Inferential statistics are used to calculate the probability that what was observed in a sample is consistent with what occurs in the unknown population.
    - We're making inferences from observed sample data to unknown population parameters.
  - As such, *inferential statistics is not about being 100% certain*.
    - We do not "prove" a hypothesis to be true or false.
    - We either find evidence against the null hypothesis, or we do not.
    - The objective is to reduce the uncertainty around the hypotheses.

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

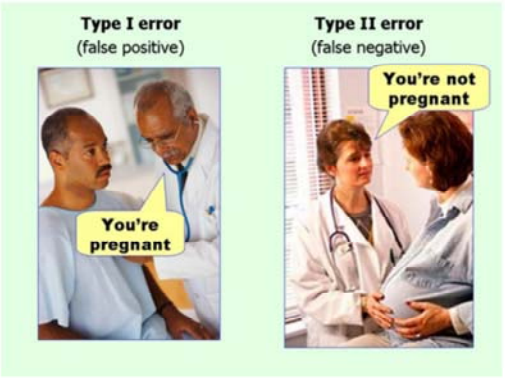
- Interpreting results – draw conclusions
  - Statistical hypothesis testing assumes no relationship/no difference (Ho) unless statistically significant evidence ($p<.05$) is found to the contrary.
  - *Study limitations and potential sources of error*

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- Type I and Type II errors

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Rejection
Fail to Reject H_0	Correct Decision	Type II Error



TheLearningCentre UNLOCK YOUR POTENTIAL

Images accessed: http://whatilearned.wikia.com/wiki/Hypothesis_Testing
http://marginalrevolution.com/marginalrevolution/2014/05/type-i-and-type-ii-errors-simplified.html

96

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

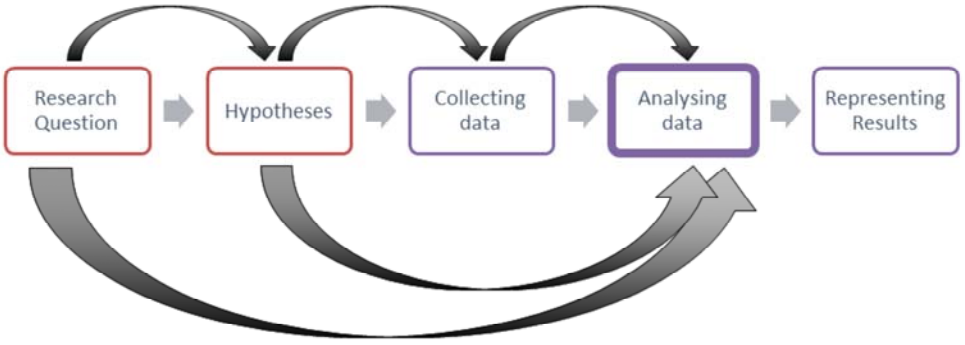
- Reporting results
  - Decision to reject (or not) the Null hypothesis
    - Supported by significance p value
  - Report within the context of study limitations/ sources of error.

6) Analysing data: Part II – Inferential Statistics

JAMES COOK UNIVERSITY AUSTRALIA

- SUMMARY
  - Inferential statistics involves hypothesis testing
    - Determine appropriate test
    - Perform test
    - Compare results to an established significance level (usually $p<0.05$)
    - Interpret results & draw conclusions.
  - Probability testing (not certainty testing)
  - Enables us to draw inferences to the population based on observed sample data

6) Analysing data: Part II – Inferential Statistics



7) Representing results

- Journal article
- Poster presentation
- Thesis
- Report

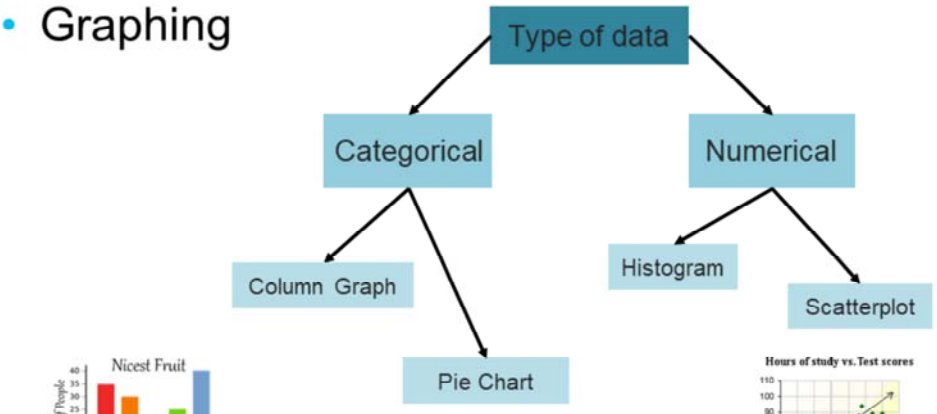
Result section

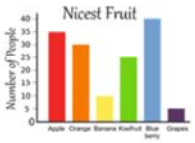
Graphs Tables

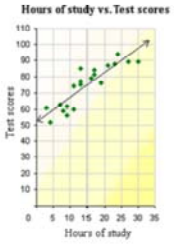
7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

- Graphing







TheLearningCentre
UNLOCK YOUR POTENTIAL

101

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

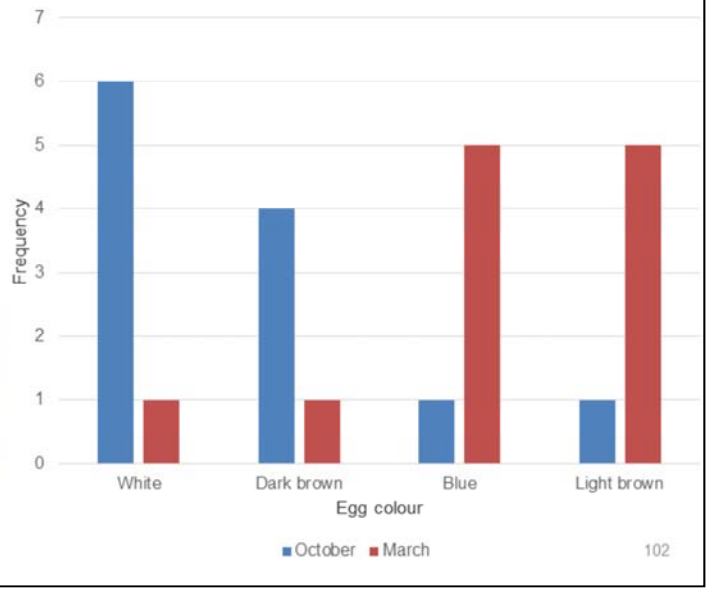
Column Graph

Eggs Laid in October

Egg Colour	Frequency
White	6
Dark brown	4
Blue	1
Light brown	1

Eggs Laid in March

Egg Colour	Frequency
White	1
Dark brown	1
Blue	5
Light brown	5



TheLearningCentre
UNLOCK YOUR POTENTIAL

102

Make sure you use axis titles and a legend

Each category has its own bar with gaps between bars

Column graphs can be combined to compare data

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

Pie Chart

- You cannot plot two variables on the same pie chart but you can compare two pie charts

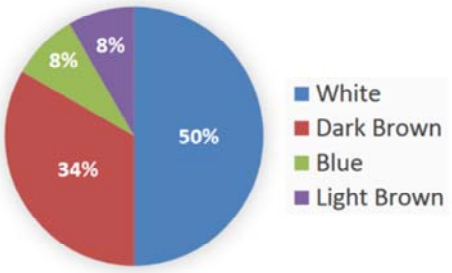


Figure 6. The colour of eggs laid during a two day period in October

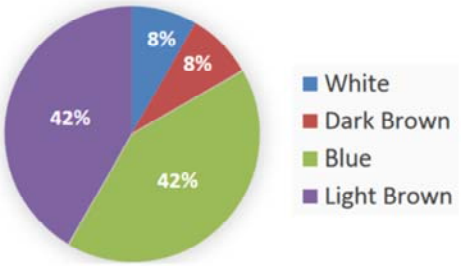


Figure 7. The colour of eggs laid during a two day period in March

103

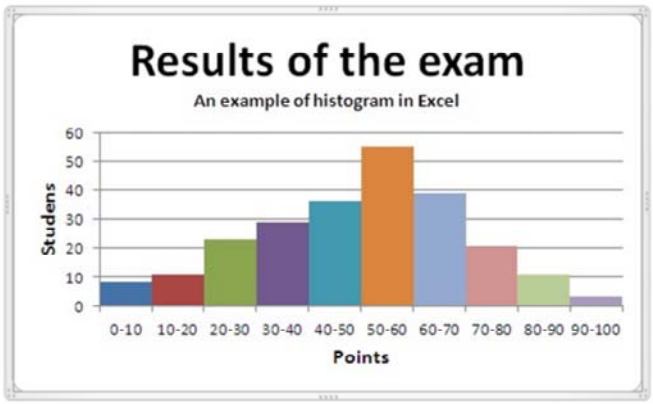
The area of the circle is proportional to the frequency

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

Histogram

- Visual display of frequencies
- Numerical variables
- Shows distribution



TheLearningCentre UNLOCK YOUR POTENTIAL

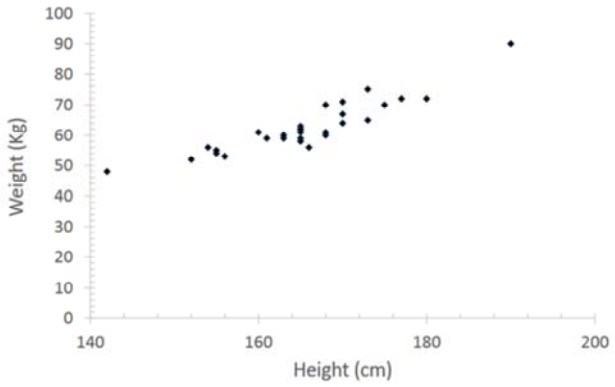
Image accessed: http://best-excel-tutorial.com/56-charts/76-histogram 104

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

Scatterplot

- Shows association between two numerical variables



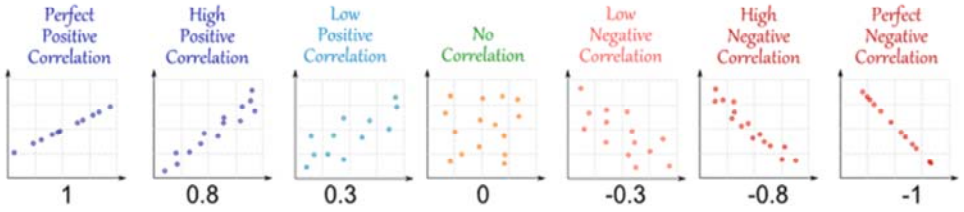
TheLearningCentre
UNLOCK YOUR POTENTIAL

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

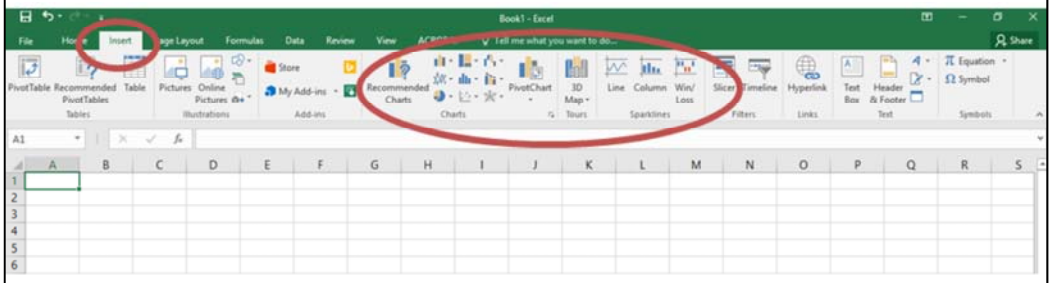
Scatterplot

- Correlations



TheLearningCentre
UNLOCK YOUR POTENTIAL

7) Representing results



- Highlight the data that you want to include in the graph
- **Lynda.com ("Excel graphing")**

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

Tables

If a graph is not suitable to represent your results, choose a table

Refer to your reference style or subject outlines for information about how to format graphs and tables

TheLearningCentre
UNLOCK YOUR POTENTIAL

Table 1

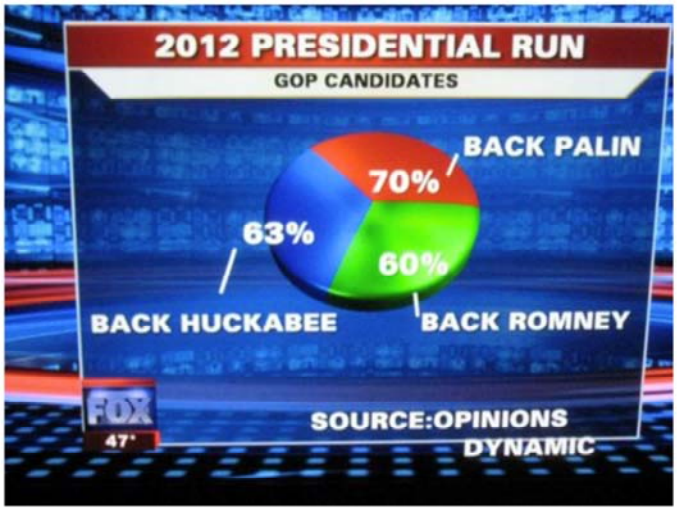
Chocolate Preferences by Employment Status

What is your current employment status?		White	Dark	Milk
Not employed	*Mean*	2.60	4.37	2.17
	SD	3.40	2.73	2.79
	N	24	28	24
Part time	*Mean*	1.87	4.71	1.15
	SD	1.95	3.15	1.38
	N	48	56	46
Full time or more	*Mean*	3.39	6.22	2.00
	SD	2.79	3.37	1.63
	N	14	16	7
Total	*Mean*	2.32	4.85	1.55
	SD	2.60	3.11	1.98
	N	86	100	77

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

No Goes



TheLearningCentre
UNLOCK YOUR POTENTIAL

109

Avoid 3 D

Use informative captions

Do not overload graphs

7) Representing results

JAMES COOK UNIVERSITY AUSTRALIA

ACTIVITY 6

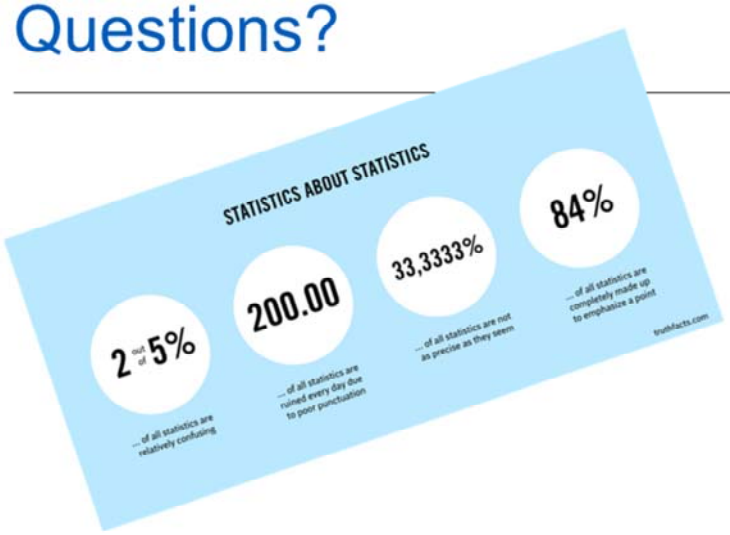
Together with the person next to you, use excel to create a

- Pie chart displaying the rainfall in 1941
- Column graph comparing the rainfall in 1941 and 1942 for all months
- Compare the average yearly rainfall
- Was there significant difference in rainfall over the 7 decades

TheLearningCentre
UNLOCK YOUR POTENTIAL

Questions?

JAMES COOK UNIVERSITY AUSTRALIA



TheLearningCentre
UNLOCK YOUR POTENTIAL

JAMES COOK UNIVERSITY AUSTRALIA

- Lecturers, tutors and supervisors
- The Learning Centre
  - https://www.jcu.edu.au/students/learning-centre/maths-and-statistics
  - Learning Advisors/Peer Advice Desk
- https://www.khanacademy.org/ KHANACADEMY
  - (good intro to chi-square test, regression, ANOVA)
- Lynda.com (structured training tutorials: e.g. Statistics foundations & applications; Excel; SPSS; R and more)
- JCU Graduate Research School 'StatsHelp' Program:
  - PhD & Research Masters
  - Professional Doctorate, or
  - Honours and coursework Masters Degree (fee basis only and pending availability).

lynda.com

TheLearningCentre UNLOCK YOUR POTENTIAL

112