Project Maths
Tionscadal Mata
Development Team

Module 1

***© Project Maths Development Team** – Draft (Version 2.0)*

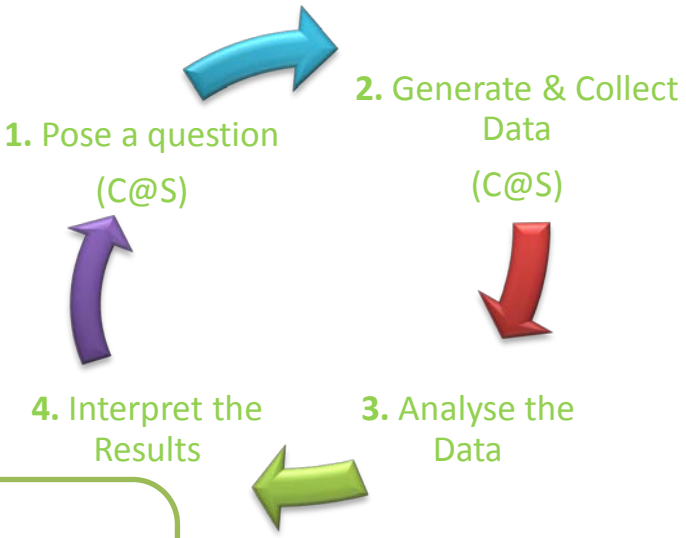
Statistics

(a) Primary sources:
- **(i)** Observational studies (JCHL, LCOL)
- **(ii)** Designed experiments (JC)

(b) Secondary sources

Sampling:
- **(i)** Random (JC)
- **(ii)** Stratified (LCHL)
- **(iii)** Cluster (LCHL)
- **(iv)** Quota (LCHL)

(a) Reliability of Data (JCHL)
(b) Summarise Data (Spreadsheets)
(c) Types of Data JC



Types of data:
Categorical/Numerical(JC)

(a) Univariate Categorical (JC)
- Pie Charts (JC)
- Bar Charts (JC)
- Line Plots (JC)

Univariate Numeric
- Histograms (JC)
- Stem and Leaf(JC)
- Back to Back (JCHL)
- Line plots (JC)

(b) Bivariate (LC)

Bivariate Numeric
- Scatter plots (LCOL)
- Correlation (LCOL)

(a) Central Tendency
- Mean (JCHL)
- Median (JC)
- Mode (JC)

(b) Spread
- Range (JCOL)
- Interquartile (JCHL)
- Standard Deviation (Calculator)

(c) Histograms
- Symmetry (LCOL)
- Skewness (LCOL)

(d) Line of best fit (LCHL)
- Correlation Coefficient
- Meaning of (LCOL)
- Calculate (LCHL)

Misuses and Misconceptions

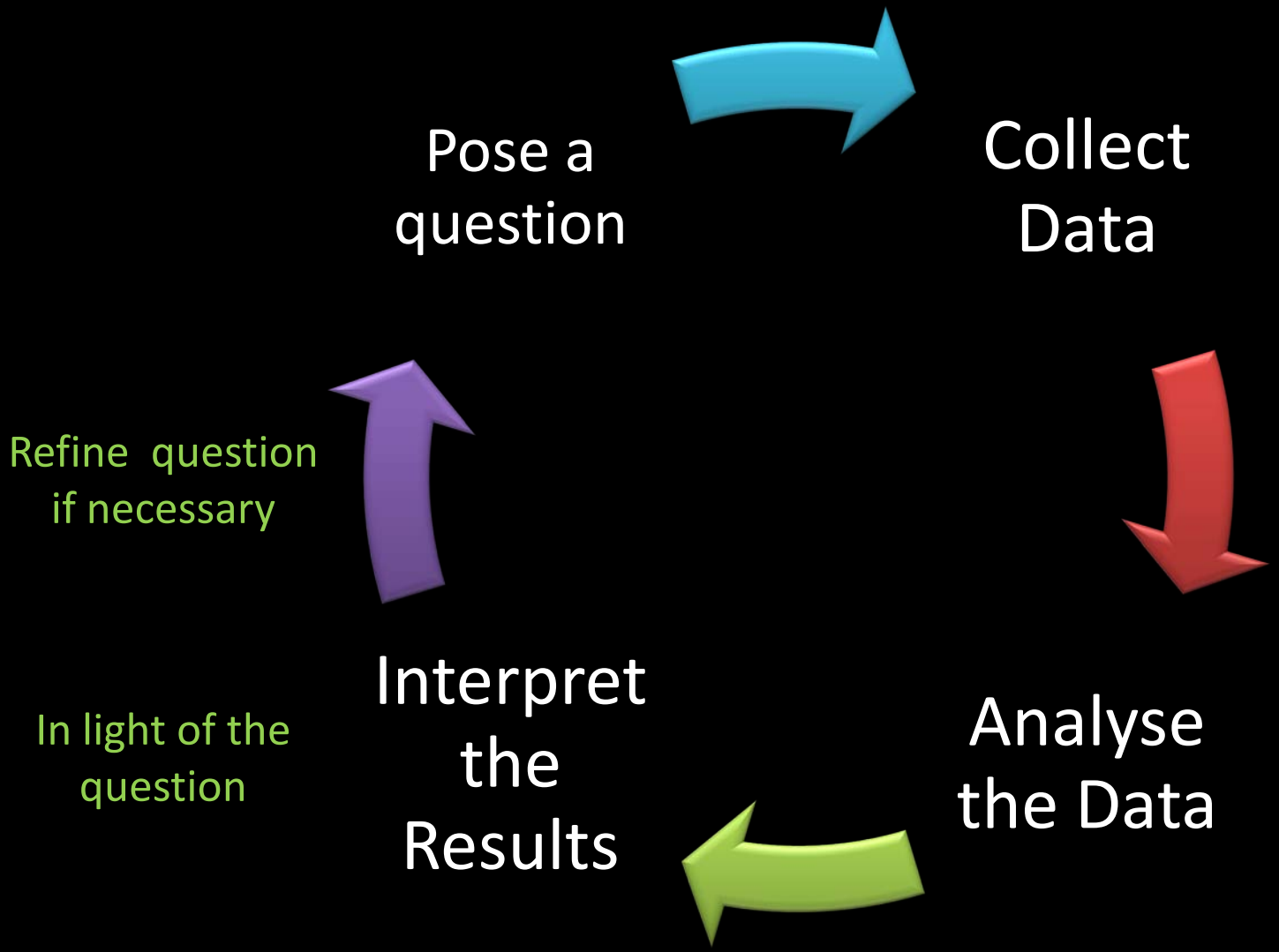
Census at School (C@S)

Students learn about:

- The use of statistics to gather information from a selection of the population with the intention of making generalisations about the whole population
- They consider situations where statistics are misused and learn to evaluate the reliability and quality of data and data sources.

(Syllabus)

The Data Handling Cycle



Producing Data

Primary Data

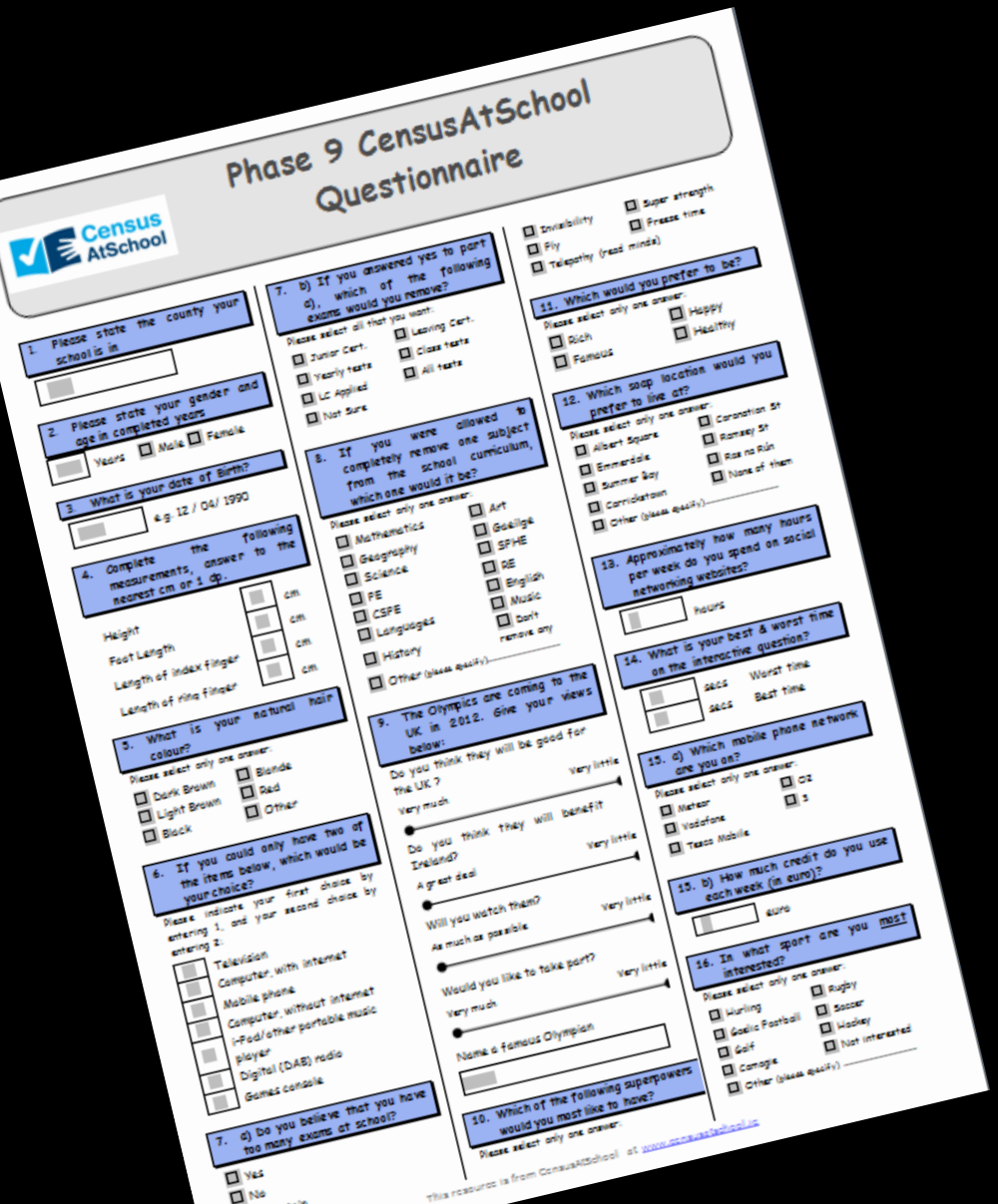
- Students collect the data themselves
- Observational studies: the researcher collects information but does not influence events e.g. surveys, epidemiological studies
- Experimental Studies: the researcher deliberately influences events and investigates the effects of the intervention, e.g. clinical trials, laboratory studies

Secondary Data

- Data collected by someone other than the user, i.e. the data already exists in books, journals, the internet etc.

"These studies always remind me of an ant colony I had as a kid!"

A Sample Survey
Census at School Questionnaire



Phase 9 CensusAtSchool Questionnaire

Census AtSchool

1. Please state the county your school is in

2. Please state your gender and age in completed years

Years ☐ Male ☐ Female

3. What is your date of Birth?

e.g. 12 / 04 / 1990

4. Complete the following measurements, answer to the nearest cm or 1 dp.

Height
Foot Length
Length of index finger
Length of ring finger

cm cm cm cm

5. What is your natural hair colour?

Please select only one answer.

☐ Dark Brown ☐ Blonde
☐ Light Brown ☐ Red
☐ Black ☐ Other

6. If you could only have two of the items below, which would be your choice?

Please indicate your first choice by entering 1, and your second choice by entering 2.

Television
Computer, with internet
Mobile phone
Computer, without internet
i-Pod/other portable music player
Digital (DAB) radio
Games console

7. a) Do you believe that you have too many exams at school?

☐ Yes
☐ No

7. b) If you answered yes to part a), which of the following exams would you remove?

Please select all that you want.

☐ Junior Cert. ☐ Leaving Cert.
☐ Yearly tests ☐ Class tests
☐ LC Applied ☐ All tests
☐ Not Sure

8. If you were allowed to completely remove one subject from the school curriculum, which one would it be?

Please select only one answer.

☐ Mathematics ☐ Art
☐ Geography ☐ Gaeilge
☐ Science ☐ SPHE
☐ PE ☐ RE
☐ CSPE ☐ English
☐ Languages ☐ Music
☐ History ☐ Don't remove any
☐ Other (please specify)

9. The Olympics are coming to the UK in 2012. Give your views below:

Do you think they will be good for the UK? Very much / Very little

Do you think they will benefit Ireland? A great deal / Very little

Will you watch them? As much as possible / Very little

Would you like to take part? Very much / Very little

Name a famous Olympian

10. Which of the following superpowers would you most like to have?

Please select only one answer.

☐ Invisibility ☐ Super strength
☐ Fly ☐ Freeze time
☐ Telepathy (read minds)

11. Which would you prefer to be?

Please select only one answer.

☐ Rich ☐ Happy
☐ Famous ☐ Healthy

12. Which soap location would you prefer to live at?

Please select only one answer.

☐ Albert Square ☐ Coronation St
☐ Emmerdale ☐ Ramsey St
☐ Summer Bay ☐ Ros na Rún
☐ Carrickstown ☐ None of them
☐ Other (please specify)

13. Approximately how many hours per week do you spend on social networking websites?

Hours

14. What is your best & worst time on the interactive question?

secs Worst time
secs Best time

15. a) Which mobile phone network are you on?

Please select only one answer.

☐ Meteor ☐ O2
☐ Vodafone ☐ 3
☐ Tesco Mobile

15. b) How much credit do you use each week (in euro)?

Euro

16. In what sport are you most interested?

Please select only one answer.

☐ Hurling ☐ Rugby
☐ Gaelic Football ☐ Soccer
☐ Golf ☐ Hockey
☐ Camogie ☐ Not interested
☐ Other (please specify)

This resource is from CensusAtSchool at www.censusatschool.ie

12. Which soap location would you prefer to live at?

Please select only one answer:

☐ Albert Square ☐ Coronation St
☐ Emmerdale ☐ Ramsey St
☐ Summer Bay ☐ Ros na Rún
☐ Carrickstown ☐ None of them
☐ Other (please specify).....

An Example of an Observational Study

Step 1: **Pose the question**

How accurate are students at estimating, to within 5 seconds, the number of seconds in a minute?

Step 2: **Collect the data**

Working in pairs (students A and B) and using a stop watch, students estimate the number of seconds in a minute. Student A signals when he/she is starting the stop watch and student B says "stop" when he/she thinks a minute has elapsed. Student B records the number of seconds estimated for a minute. The students then switch roles so that this time student A estimates and B operates the stop clock.

Step 3: **Analyse the data**

Estimated times from all the groups are recorded and a stem and leaf plot is produced for the whole class.

Step 4: **Interpret the result**

Are there any/many outliers? Find the mode and median from the stem plot and calculate the mean.

Are these values close to 60 seconds?

Answer the original question.

Extension Question:

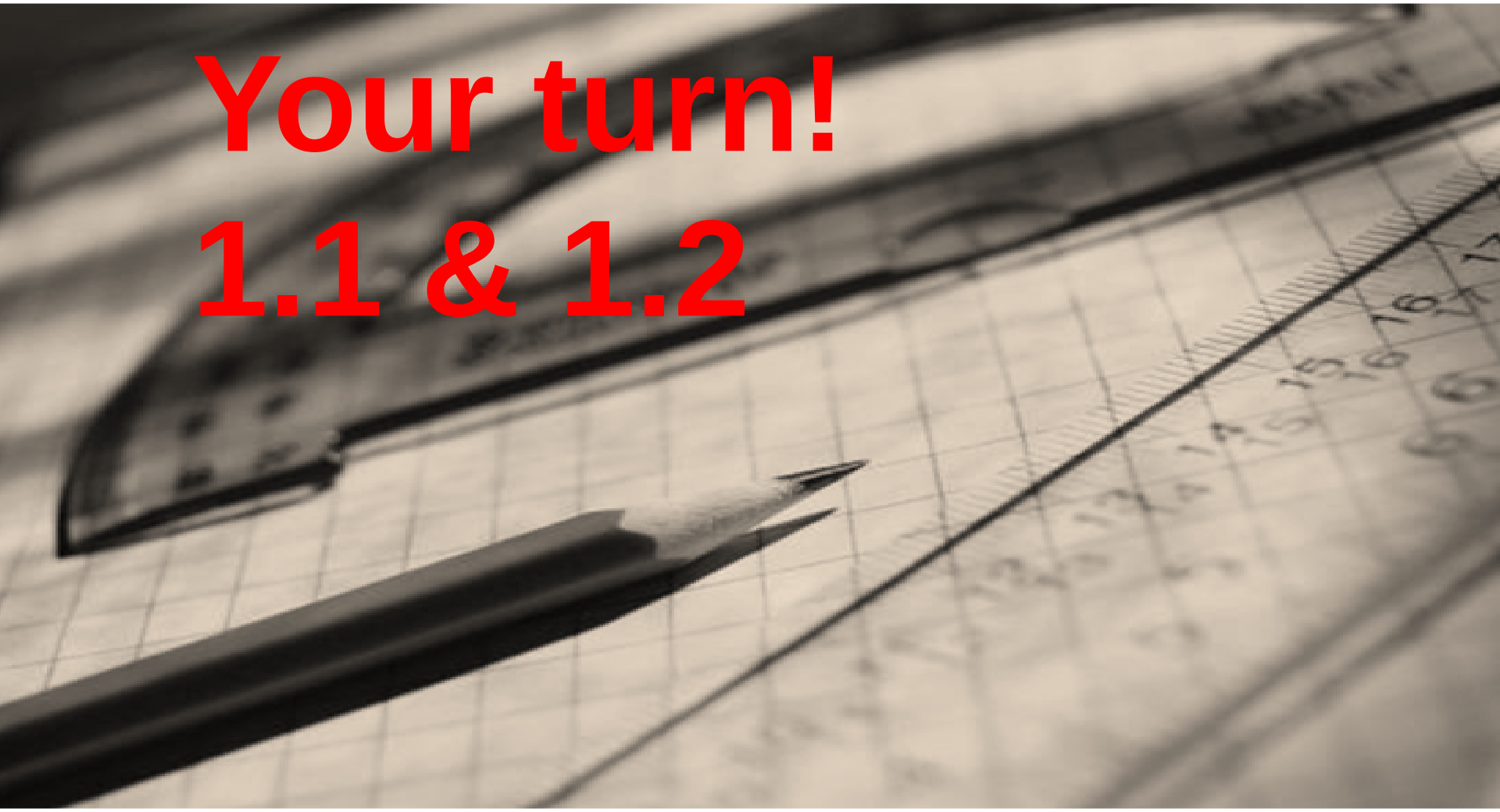
Do students get better at estimating the number of seconds in a minute with practice?

What happens to the number of outliers which each successive trial?

What number are most of the data points clustered around etc.?

How Reliable is Secondary Data?

Who carried out the survey?
What was the population?
How was the sample selected?
How large was the sample?
What was the response rate?
How were the subjects contacted?
When was the survey conducted?
What were the exact questions asked?


Your turn!
1.1 & 1.2

1.1

Which of the following statements is correct regarding observational studies?

A researcher can observe but not control the explanatory variables. ✓

A researcher can define but not observe the explanatory variables. ✗

A researcher can minimise but not eliminate the explanatory variables. ✗

A researcher can control but not observe the explanatory variables. ✗

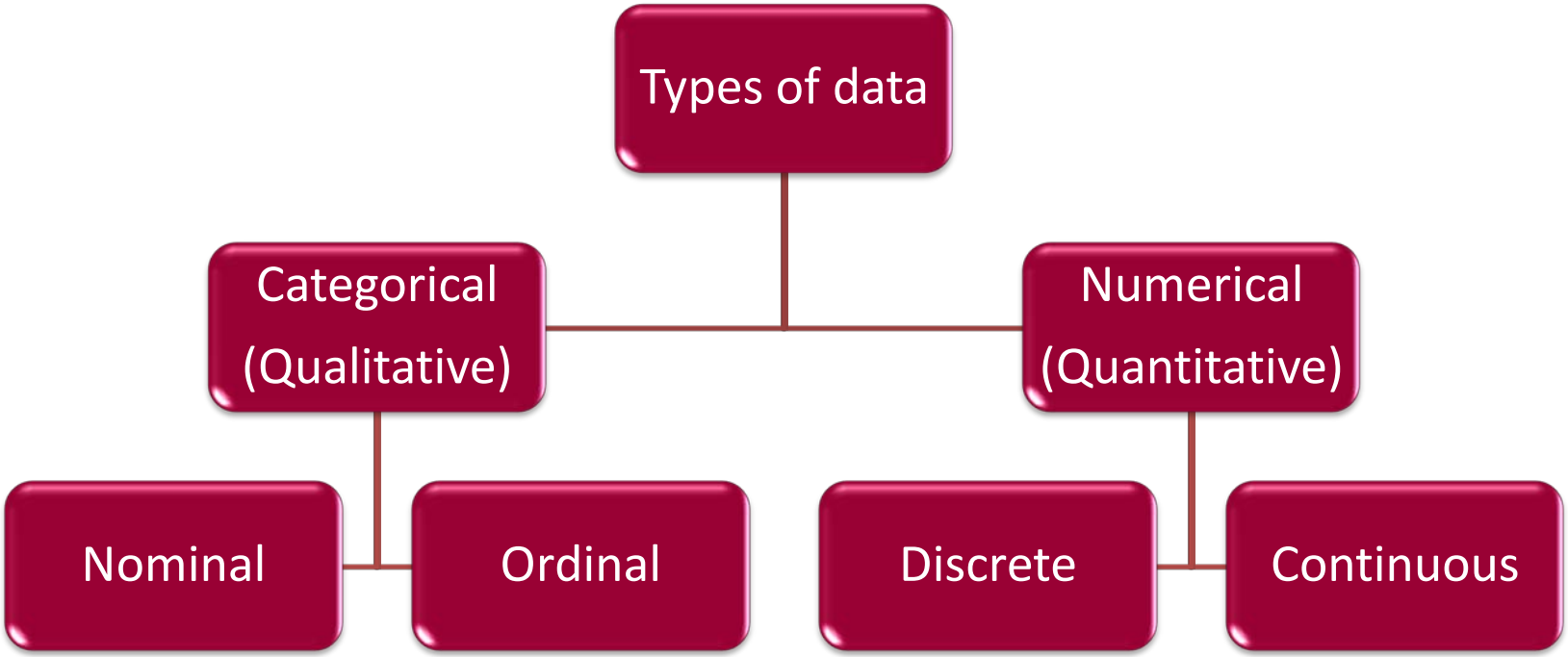
© Project Maths Development Team – Draft (Version 2.0)

1.2

Some people believe that exercise raises the body's metabolic rate for as long as 12 to 24 hours and thus enables us to continue to burn off fat after we end our workout. In a study of this effect, subjects were asked to walk briskly on a treadmill for several hours. Their metabolic rate was measured before, immediately after and 12 hours after the exercise. Was this study an experiment? Why or why not? What are the explanatory and response variables?

Solution

Data Types


Types of data
Categorical (Qualitative)
Numerical (Quantitative)
Nominal
Ordinal
Discrete
Continuous

Univariate Data

(1) Involves a **single variable i.e.** we look at one item of data at a time from each subject e.g. height

(2) Not dealing with relationships between variables

(3) The major purpose of Univariate analysis is to describe

Sample question:

How many of the students in the class are female?

A Bar Chart

Note: A bar chart describes categorical data, and has gaps, whereas a histogram describes continuous data and hence has no gaps.

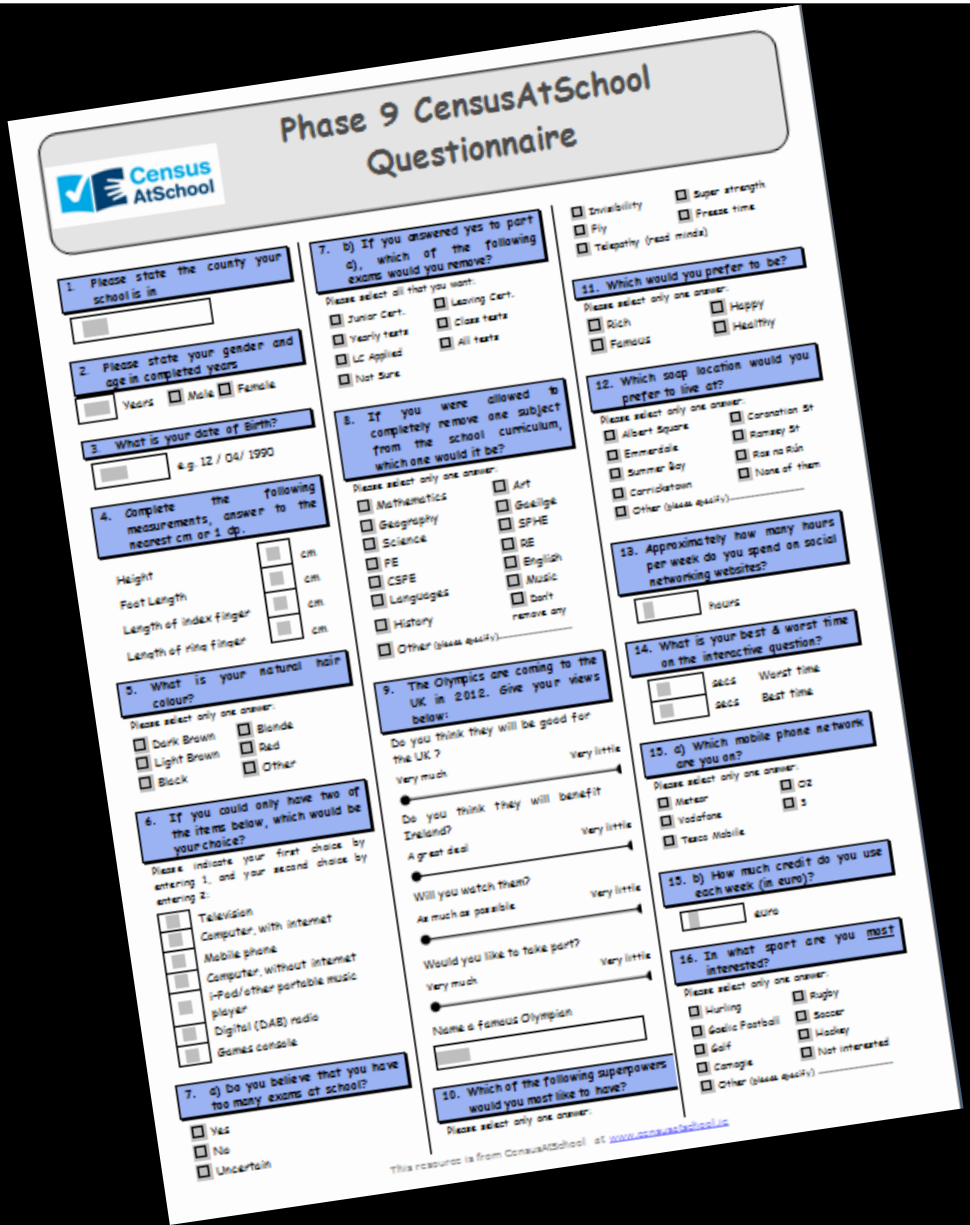
Categorical

Type	Nominal	Ordinal
Description	Can be identified by particular names or categories and cannot be organised according to any natural order	Data which looks like numbers but are really just labels, they can be identified by categories which can be ordered in some way
Examples	**Gender:** Male or female **Hair Colour:** black, blonde etc., **Favourite Sport:** Soccer, rugby	ISBN Numbers, Visa card no., **Watching TV**: Never, Rarely, Sometimes, A lot
Suitable Graphical Representation	Bar chart, Line plot, Pie chart	Bar chart, Line plot, Pie chart

Numerical

Type	Discrete	Continuous
Description	Data can only have a finite number of values	Data can assume an infinite number of values between any two given values. Students height may be 1.4325 m
Examples	No. of peas in a pod, age in years (as opposed to age)	Height, arm span, foot length
Suitable Graphical Representation	Bar chart, Pie chart, Line plot, Stem plot	Histogram

Types of Data
Census at School Questionnaire



Phase 9 CensusAtSchool Questionnaire

Census AtSchool

1. Please state the county your school is in

2. Please state your gender and age in completed years
Years ☐ Male ☐ Female

3. What is your date of Birth?
e.g. 12 / 04/ 1990

4. Complete the following measurements, answer to the nearest cm or 1 dp.

Height	cm
Foot Length	cm
Length of index finger	cm
Length of ring finger	cm

5. What is your natural hair colour?
Please select only one answer.
☐ Dark Brown ☐ Blonde
☐ Light Brown ☐ Red
☐ Black ☐ Other

6. If you could only have two of the items below, which would be your choice?
Please indicate your first choice by entering 1, and your second choice by entering 2:
Television
Computer, with internet
Mobile phone
Computer, without internet
i-Pod/other portable music player
Digital (DAB) radio
Games console

7. a) Do you believe that you have too many exams at school?
☐ Yes
☐ No
☐ Uncertain

7. b) If you answered yes to part a), which of the following exams would you remove?
Please select all that you want:
☐ Junior Cert. ☐ Leaving Cert.
☐ Yearly tests ☐ Class tests
☐ LC Applied ☐ All tests
☐ Not Sure

8. If you were allowed to completely remove one subject from the school curriculum, which one would it be?
Please select only one answer:
☐ Mathematics ☐ Art
☐ Geography ☐ Gaeilge
☐ Science ☐ SPHE
☐ PE ☐ RE
☐ CSPE ☐ English
☐ Languages ☐ Music
☐ History ☐ Don't remove any
☐ Other (please specify)

9. The Olympics are coming to the UK in 2012. Give your views below:
Do you think they will be good for the UK ?
Very much — Very little
Do you think they will benefit Ireland?
A great deal — Very little
Will you watch them?
As much as possible — Very little
Would you like to take part?
Very much — Very little
Name a famous Olympian

10. Which of the following superpowers would you most like to have?
Please select only one answer:
☐ Invisibility ☐ Super strength
☐ Fly ☐ Freeze time
☐ Telepathy (read minds)

11. Which would you prefer to be?
Please select only one answer:
☐ Rich ☐ Happy
☐ Famous ☐ Healthy

12. Which soap location would you prefer to live at?
Please select only one answer:
☐ Albert Square ☐ Coronation St
☐ Emmerdale ☐ Ramsay St
☐ Summer Bay ☐ Fair na Rún
☐ Corriekstown ☐ None of them
☐ Other (please specify)

13. Approximately how many hours per week do you spend on social networking websites?
hours

14. What is your best & worst time on the interactive question?
secs Worst time
secs Best time

15. a) Which mobile phone network are you on?
Please select only one answer:
☐ Meteor ☐ O2
☐ Vodafone ☐ 3
☐ Tesco Mobile

15. b) How much credit do you use each week (in euro)?
euro

16. In what sport are you most interested?
Please select only one answer:
☐ Hurling ☐ Rugby
☐ Gaelic Football ☐ Soccer
☐ Golf ☐ Hockey
☐ Camogie ☐ Not interested
☐ Other (please specify)

This resource is from CensusAtSchool at www.censusatschool.ie

Video – Data Types

The notes accompanying the video can be downloaded from

www.nce-mstl.ie

under Resources, Secondary Level, Mathematics, Summer Course – Statistics and Probability

Duration: 00:02:34

Your turn!
1.3

What type of data is generated by each of the questions in the Census at School Survey?

Solution

Let's Define Some Terms

A subset of the population.

- **Population:** The entire group of subjects about which information is required
- **Sample:** Any subset of a population e.g. a representative subset of students from the school
- **Variable:** We measure its value for each person and it varies from person to person e.g. the height of an individual or their favourite sport
- **Parameter:** Some value we are interested in calculating for the population
- **Statistic:** Some value we are interested in calculating for the sample

Your turn!
1.4

1.4

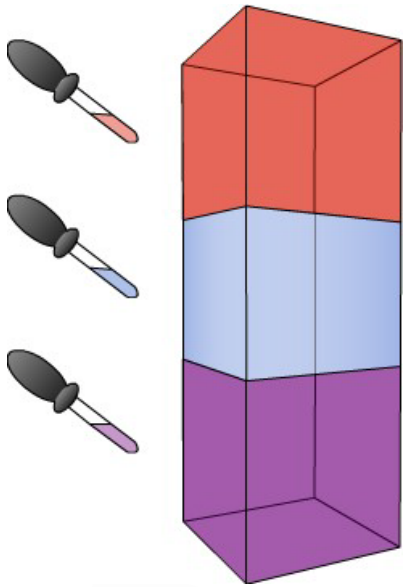
The Gardai Siochana wants to know how Dublin inner city residents feel about the police service. A questionnaire with several questions about the police is prepared. A sample of 300 mailing addresses in inner city areas is chosen, and a Garda is sent to each address to administer the questionnaire to an adult living there. Identify the population, variables measured and the sample. In addition, describe the potential bias.

[NCE – MSTL, Q2, Pg 16]

Solution

Types of Sampling

- Simple Random Sampling
- Stratified Random Sampling
- Cluster Sampling
- Quota Sampling



Simple Random Sample

One way of collecting data is to use a Sample. Whenever you need to take a sample the sample will need to be a Random Sample which is Representative of the population.

Example:
A new business with 100 employees wants to know whether staff would like to have childcare facilities on site.

An estimate could be made by asking a sample of 20 employees if they would use the childcare facilities and multiplying the number who say yes by 5.

If we do this, we have to decide which people to ask.

Biased Samples

When we are taking samples it is very important to avoid Bias.

Suppose we want to estimate how many students watch the X-Factor in a school with 1000 students.

Suppose we take a random sample of 50 and ask if they watch the X-Factor

. . . and all in the sample happen to be girls.

(Very unlikely but possible)

If girls are more, or less, likely to watch the X-Factor than boys we would have a biased sample.

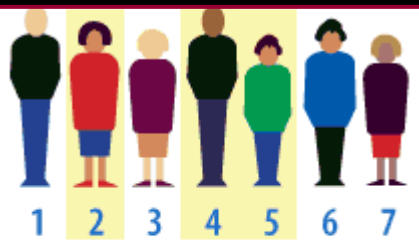
Our results could be Unreliable

So, we need to avoid bias.

Video – Bias

Duration: 00:05:52

Random Sample



Assign Numbers, Auto-Generate Random Selections

Random does NOT mean that we can just pick anyone for the sample.

To get a Random Sample of 20 people we could give each person a number from 1 to 100 and then select 20 numbers using a random method.

One Random method is to write the 100 numbers on separate slips, put them in a bag, shake them, and take 20 of them out without looking.

A better random method for large samples is to use the Random Number Generator found on your calculator.

Generating Random Numbers using a Calculator

The button might say “RANDOM” (SHARP).
Other makes may have a button “Ran” or “Ran#” or “RanInt”.
Whichever you have, selecting and pressing ENTER repeatedly gives random numbers.

Generate a Random Number between 0 and 99.

Sharp EL–520W & EL–W531

100 | 2nd F | 7 | 0 | Enter

Casio fx–83ES

Shift | Mode | 6 | 0

100 | Shift | . | =

N.B. Calculator should be in LINE IO mode. Shift mode 2

Stratified Random Sample

Example:

Suppose there are 500 girls and 500 boys.

Decide with the person beside you how you could avoid gender bias in taking a sample of 50.

Answer:

Take 2 random samples, one of 25 boys and one of 25 girls, and then combine them.

However, we are unlikely to have exactly equal numbers of boys and girls.

Can you see what to do if the school has 560 boys and 440 girls and we need a sample of 50 ?

Problem

How many of each of the 3 types of computer component should be taken in a sample of 100 categorised by type of component?

Type	A	B	C	Total
Number	300	260	40	600

Solution:

The total number of components $= 600$

Component A: $\frac{300}{600} \times 100 = 50$

Component B: $\frac{260}{600} \times 100 = 43.3$ [Round to 43]

Component C: $100 - 50 - 43 = 7$

We need 50, 43 and 7 respectively.

Cluster & Quota Sampling

Cluster

- Splitting the population into similar parts or clusters can make sampling more practical.
- Then we could simply select one or a few clusters at random and perform a **census** within each of them.

Quota

- Non probability sampling method
- Example: Opinion Polls
  - 1000 – (2000) in all 43 constituencies
  - Split by gender, age, rural, urban, etc.
  - Not truly random, not equal chance of being selected as interviewer has been told what to get

Video – Quota Sampling

Duration: 00:01:13

Your turn!
1.5

1.5

We need to survey a random sample of the 300 passengers on a flight from San Francisco to Tokyo. Name each sampling method described below.

(a) From the boarding list, randomly choose 5 people flying first class and 25 of the other passengers.

(b) Randomly generate 30 seat numbers and survey the passengers who sit there.

(c) Randomly select a seat position (right centre, right window, right aisle etc.) and survey all passengers sitting in those seats.

(d) From the boarding list, select 30 passengers of which 15 are male, 15 females all in the 40 – 50 age bracket.

Solution

(a) Stratified

(b) Simple

(c) Cluster

(d) Quota

© Project Maths Development Team – Draft (Version 2.0)

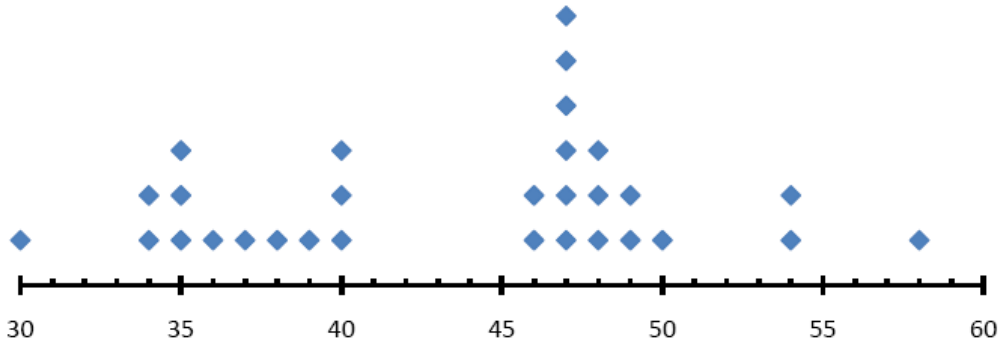
Representing Data Graphically – Line Plots (Univariate)

Example: Suppose thirty people live in an apartment building. These are their ages:

58, 30, 37, 36, 34, 49, 35, 40, 47, 47, 39, 54, 47, 48, 54,
50, 35, 40, 38, 47, 48, 34, 40, 46, 49, 47, 35, 48, 47, 46

Represent this data using a line plot.

Solution:



Note:

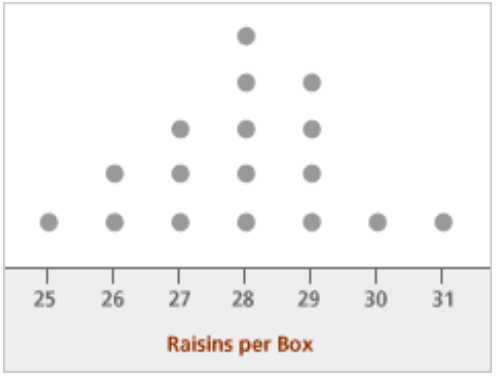
Clusters are isolated groups of points, such as the ages of 46 through 50.

Gaps are large spaces between points, such as 41 and 45.

Outliers are items of data which lie far away from the overall pattern of the rest of the data, such as the data for ages 30 and 58.

Your turn!
1.6

Students were investigating the number of raisins contained in individual mini boxes of Sun-Maid raisins.



(i) How many boxes of raisins did they survey?

(ii) What was the modal number of raisins per box?

(iii) What is the median number of raisins per box? Explain how you found this answer.

[NCCA – Student Resources, Q5, LC, pg. 83]

Solution

Representing Data Graphically, Stem and Leaf Plots

The ages of the 30 members of an aerobics class are:

19	22	31	17	8	12	23	47	53	47	19	46	38	59	47
52	21	58	54	26	32	47	55	62	64	36	37	43	15	51

These are presented as a stem and leaf diagram by using the first digit as **stem**, and the second digit as the **leaf**:

19 is written 1|9
22 is written 2|2
31 is written 3|1

17 is then 1|9 7

and 8 is 0|8

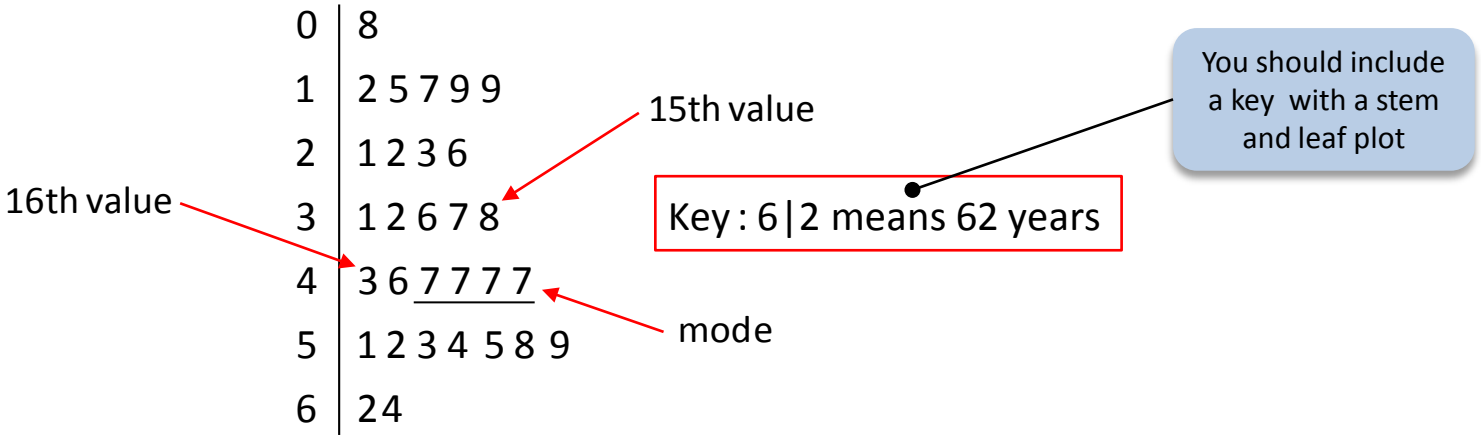
0	8
1	9 7
2	2
3	1
4	
5	
6	

So the complete diagram is:

0	8
1	9 7 2 9 5
2	2 3 1 6
3	1 8 2 6 7
4	7 7 6 7 7 3
5	3 9 28 4 5 1
6	24

A Stem and Leaf plot is like a histogram but it shows the individual values

Now put the leaves in order:



From this you can pick out the mode and identify the median.

Mode $= 47$

The median is the $\frac{30+1}{2} = 15\frac{1}{2}$th value $= 40.5$

The average of the 15th and 16th values is $\frac{38+43}{2} = 40.5$

15	7	7	9					
16	0	0	2	2	2	2	3	4
16	5	5	5	6	7	8		
17	0	1	1	1	2	2	3	4
17	5	7						
18	0	1	2					

© Project Maths Development Team – Draft (Version 2.0)

© Project Maths Development Team – Draft (Version 2.0)

15	7	7	9					
16	0	0	2	2	2	2	3	4
16	5	5	5	6	7	8		
17	0	1	1	1	2	2	3	4
17	5	7						
18	0	1	2					

Back to Back Stem Plot Example

Sample Paper, OL Q9

The students in a Leaving Certificate class decided to investigate their heights. They measured the height of each student, in centimetres. The heights of the boys and the girls in the class are given below:

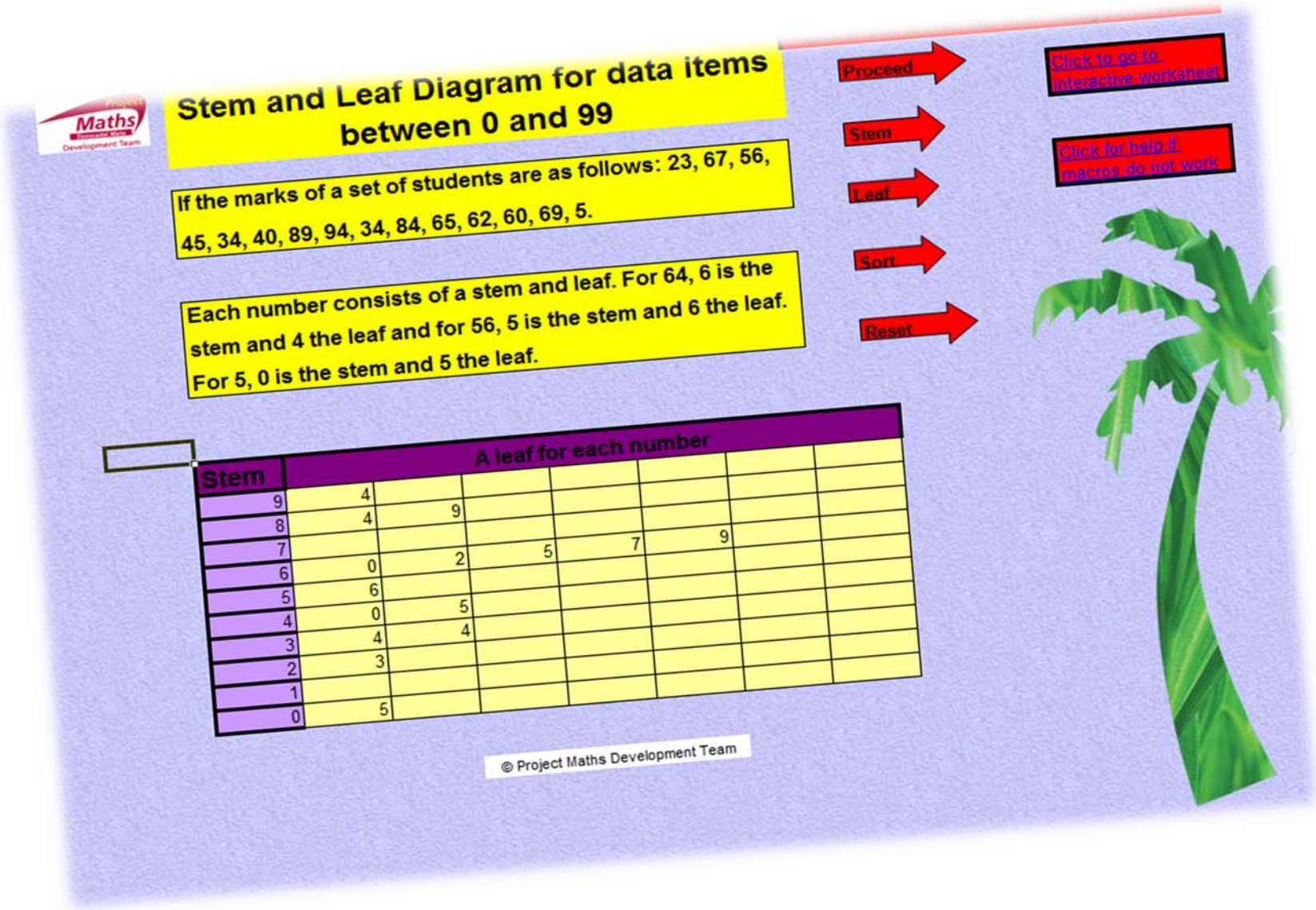
Boys						Girls					
173	180	174	175	178	176	167	161	160	157	164	172
180	171	170	187	176	166	168	149	161	167	167	171

(a) Construct a back-to-back stem and leaf plot of the above data.

(b) State one difference and one similarity between the two distributions.

Solution

(a) **(b)**


Stem and Leaf Diagram for data items between 0 and 99
If the marks of a set of students are as follows: 23, 67, 56, 45, 34, 40, 89, 94, 34, 84, 65, 62, 60, 69, 5.
Each number consists of a stem and leaf. For 64, 6 is the stem and 4 the leaf and for 56, 5 is the stem and 6 the leaf. For 5, 0 is the stem and 5 the leaf.
Proceed
Stem
Leaf
Sort
Reset
Click to go to interactive worksheet
Click for help if macros do not work
Stem
A leaf for each number
9 4
8 4 9
7 5 7 9
6 0 2
5 6
4 0 5
3 4 4
2 3
1
0 5
© Project Maths Development Team

Advantages of Stem Plots

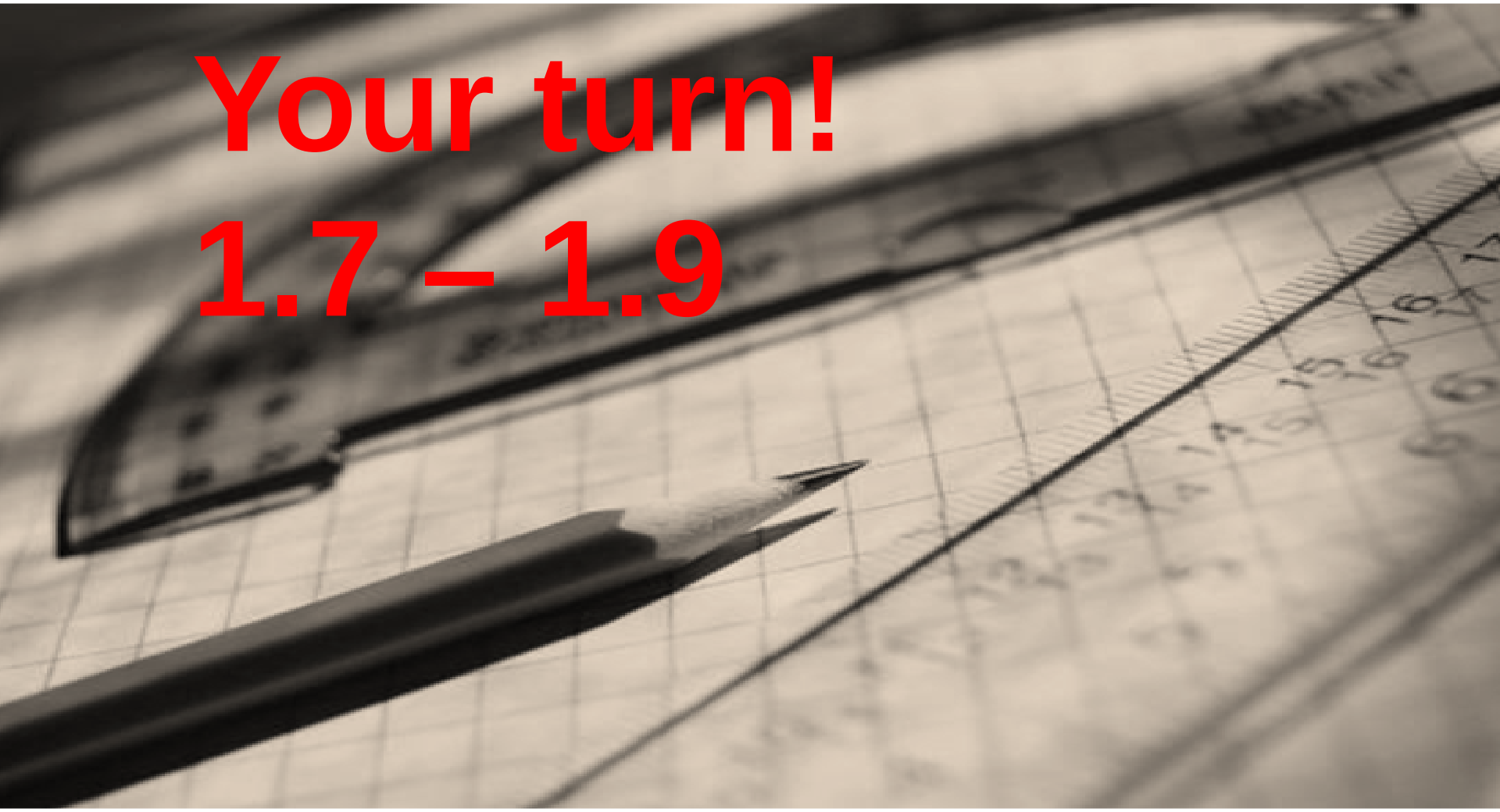
- A Stem Plot displays each separate data value
- They give a quick clear picture of a distribution and make it easy to identify clusters of data from the lengths of the branches
- A Stem and Leaf Plot shows how wide a range of values the data cover, where the values are concentrated and whether the data has any symmetry
- As the values on the branches are ordered it is very easy to pick out the median, quartiles, maximum and minimum values and to identify any outliers
- They are easily created by hand
- Two data sets can be compared using a Back-to-Back Stem Plot
- Both discrete and continuous data sets can be displayed

Disadvantages of Stem Plots

- They cannot display categorical type data

- Small data sets with a large range can be difficult to display on the stem plot without rounding e.g. 432, 507, 534, 581, 609, 626, 671, 712, 719

  [These data points could be displayed with the hundreds digits as the stem and the tens as the leaves.]


Your turn!
1.7 – 1.9

1.7

The amounts of pocket money given to 30, 5th year students per month in € are as follows:

34, 35, 32, 33, 32, 35, 34, 31, 28, 30, 31, 30, 35, 32, 45, 41,
42, 41, 46, 35, 35, 36, 36, 32, 34, 35, 33, 21, 33 & 51.

(a) Represent this data by a Stem and Leaf Plot.
(b) Explain why this type of Data is suitable to be represented by a Stem and Leaf Plot.
(c) How many students received €30 per month as pocket money?
(d) What was the modal amount of pocket money per month?
(e) What was the median amount of pocket money?

Solution

(a)

(b)

(c)

(d)

(e)

© Project Maths Development Team – Draft (Version 2.0)

1.8

The stem and leaf plot shows the time taken for 15 students to walk to school.

Key: 1|6 represents 16 mins.

(a) Find the median time taken.

(b) Find the mode

(c) Find the range

(d) What was the fastest time taken?

0	5 7 7 8
1	0 0 2 5 6 6 8
2	5 5 5
3	8

Solution

1.9

The stem and leaf plot shows the time taken for 16 students to walk to school.

Key: 2|5 represents 25 mins.

(a) Find the median time taken?

(b) Find the lower quartile.

(c) Find the upper quartile and hence the interquartile range.

0	5 7 7 8
1	0 0 2 5 6 6 8
2	5 5 5
3	8 9

Solution

Notes