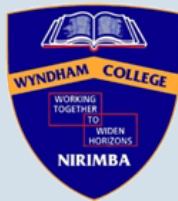


## STD 1: Statistical Analysis (Std 1), S1 Data Analysis (Y11)

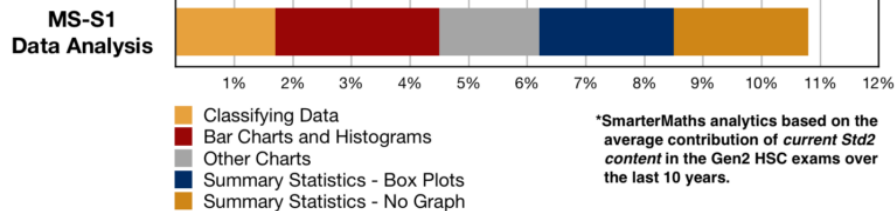
### Classifying Data (Std 1)

Teacher: Kirtana Hariharan

Exam Equivalent Time: 33 minutes (based on HSC allocation of 1.5 minutes approx. per mark)



### General 2 Exam Contribution History S1 Data Analysis



### IMPORTANT FEATURES AND TIPS FROM 2UG EXAM HISTORY

- *MS-S1 Data Analysis* was a major contributor to the old Gen2 course, contributing an average of 10.8% per exam over the past decade (note past allocations are no guarantee of future contributions to the Standard 1 exam, but can nonetheless cast light on a topic's likely importance).
- This analysis looks at the sub-topic *Classifying Data* (1.7%).

### ANALYSIS - What to Expect and Common pitfalls

- *Classifying Data* is a sneaky important sub-topic for two reasons. Firstly, it has been tested in eight exams of the last decade (only omitted in 2010 and 2014). Secondly, it has surprisingly caused difficulty with sub-50% mean marks on 3 separate occasions.
- Stratified sampling and categorical vs numerical variables (and more importantly, their respective sub-categories) deserve close attention.

## Questions

### 1. Data, 2UG 2013 HSC 8 MC

⚡ RAP Data - Bottom 15%: School result (95%) was 3% above state average (92%)

A high school has 100 students in each year group, Year 7 to Year 12. A survey is to be conducted to determine the average number of text messages sent per month by students at the school.

Which of the following would provide the most representative sample for this survey?

- (A) All Year 7 students
- (B) All physics students in Year 11 and 12
- (C) 20 students chosen at random from each year group
- (D) 120 students chosen at random from the school roll

### 2. Data, 2UG 2012 HSC 2 MC

Handmade chocolates are checked for size and shape. Every 30th chocolate is sampled.

Which term best describes this type of sampling?

- (A) Census
- (B) Random
- (C) Stratified
- (D) Systematic


### 3. Data, 2UG 2009 HSC 3 MC

The eye colours of a sample of children were recorded.

When analysing this data, which of the following could be found?

- (A) Mean
- (B) Median
- (C) Mode
- (D) Range

#### 4. Data, 2UG 2009 HSC 5 MC

 RAP Data - Bottom 18%: School result (87%) was 5% above state average (82%)

Jamie wants to know how many songs were downloaded legally from the internet in the last 12 months by people aged 18–25 years. He has decided to conduct a statistical inquiry.

After he collects the data, which of the following shows the best order for the steps he should take with the data to complete his inquiry?

- (A) Display, organise, conclude, analyse
  - (B) Organise, display, conclude, analyse
  - (C) Display, organise, analyse, conclude
  - (D) Organise, display, analyse, conclude
- 

#### 5. Data, STD2 S1 SM-Bank 1 MC


A survey asked the following question for students born in Australia:

"Which State or Territory were you born in?"

How would the responses be classified?

- A. Categorical, ordinal
  - B. Categorical, nominal
  - C. Numerical, discrete
  - D. Numerical, continuous
- 

#### 6. Data, 2UG 2015 HSC 4 MC

 RAP Data - Bottom 6%: School result (69%) was equal to state average (69%)

On a school report, a student's record of completing homework is graded using the following codes.

- C = consistently
- U = usually
- S = sometimes
- R = rarely
- N = never

What type of data is this?

- (A) Categorical, ordinal
  - (B) Categorical, nominal
  - (C) Numerical, continuous
  - (D) Numerical, discrete
- 

#### 7. Data, 2UG 2017 HSC 4 MC

A factory's quality control department has tested every 50th item produced for possible defects.

What type of sampling has been used?

- A. Random
  - B. Stratified
  - C. Systematic
  - D. Numerical
- 

#### 8. Data, 2UG 2018 HSC 3 MC

A survey asked the following question.

'How many brothers do you have?'

How would the responses be classified?

- A. Categorical, ordinal
  - B. Categorical, nominal
  - C. Numerical, discrete
  - D. Numerical, continuous
-

### 9. Data, 2UG 2007 HSC 17 MC

Ms Wigginson decided to survey a sample of 10% of the students at her school.

The school enrolment is shown in the table.

Year	7	8	9	10	11	12	Total
Number of students	225	232	233	230	150	130	1200

She surveyed the same number of students in each year group.

How would the numbers of students surveyed in Year 10 and Year 11 have changed if Ms Wigginson had chosen to use a stratified sample based on year groups?

- (A) Increased in both Year 10 and Year 11
- (B) Decreased in both Year 10 and Year 11
- (C) Increased in Year 10 and decreased in Year 11
- (D) Decreased in Year 10 and increased in Year 11

### 10. Data, 2UG 2016 HSC 7 MC

Which set of data is classified as categorical and nominal?

- (A) blue, green, yellow
- (B) small, medium, large
- (C) 5.2 cm, 6 cm, 7.21 cm
- (D) 4 people, 5 people, 9 people

### 11. Data, 2UG 2009 HSC 24c

⚡ RAP Data - Bottom 17%: School result (72%) was 4% above state average (68%)

The Australian Bureau of Statistics provides the NSW government with data on the age of residents living in different areas across the state. After analysing this data, the government makes decisions relating to the provision of services or facilities.

Give an example of a possible decision the government might make and describe how the data might justify this decision. (2 marks)

### 12. Data, 2UG 2012 HSC 26d

Greg needs to conduct a statistical inquiry into how much time people aged 18–25 years have spent accessing social media websites in the last two weeks. He has decided to survey a sample of students from his university.

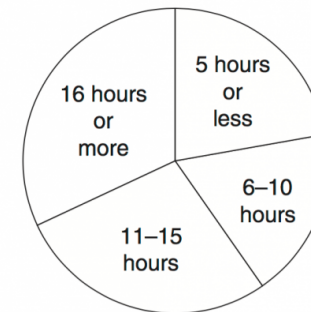
The process of statistical inquiry includes the following steps, which are NOT in order.

- A Writing a report
- B Posing questions
- C Organising data
- D Analysing data and drawing conclusions
- E Collecting data
- F Summarising and displaying data

- (i) Using the letters A, B, C, D, E and F, list the steps in the most appropriate order for Greg to conduct his statistical inquiry. (2 marks)
- (ii) Greg conducts his statistical inquiry.

At which step in the process would he have drawn this graph? (1 mark)

#### Time spent accessing social media websites (in hours)



### 13. Data, 2UG 2006 HSC 23c

Vicki wants to investigate the number of hours spent on homework by students at her high school.

- (i) Briefly describe a valid method of randomly selecting 200 students for a sample. (1 mark)
- (ii) Vicki chooses her sample and asks each student how many hours (to the nearest hour) they usually spend on homework during one week. The responses are shown in the frequency table.

<i>Number of hours spent on homework in a week</i>	<i>Frequency</i>
0 to 4	69
5 to 9	72
10 to 14	38
15 to 19	21

What is the mean amount of time spent on homework? (2 marks)

---

### 14. Data, 2UG 2011 HSC 25a

**⚡ Part iv: RAP Data - Bottom 24%: School result (25%) was 7% above state average (18%)**

A study on the mobile phone usage of NSW high school students is to be conducted.

Data is to be gathered using a questionnaire.

The questionnaire begins with the three questions shown.

Q1: Do you own a mobile phone? Yes <input type="checkbox"/> No <input type="checkbox"/>
Q2: Which phone company do you use? .....
Q3: Do you use pre-paid or a plan? Pre-paid <input type="checkbox"/> Plan <input type="checkbox"/>

- (i) Classify the type of data that will be collected in Q2 of the questionnaire. (1 mark)
- (ii) Write a suitable question for this questionnaire that would provide quantitative data. (1 mark)
- (iii) An initial study is to be conducted using a stratified sample.  
Describe a method that could be used to obtain a representative stratified sample. (1 mark)
- (iv) Who should be surveyed if it is decided to use a census for the study? (1 mark)

## Worked Solutions

### 1. Data, 2UG 2013 HSC 8 MC

The best sample would have an equal amount of people in each year randomly selected.

⇒ *C*

---

### 2. Data, 2UG 2012 HSC 2 MC

Systematic Sampling

⇒ *D*

---

### 3. Data, 2UG 2009 HSC 3 MC

Eye colour is categorical data

∴ Only the mode can be found

⇒ *C*

---

### 4. Data, 2UG 2009 HSC 5 MC

Process of statistical enquiry requirements

⇒ *D*

---

### 5. Data, STD2 S1 SM-Bank 1 MC

The data is categorical (not numerical) since the name of a State is required.

This data cannot be ordered.

⇒ *B*

---

## Worked Solutions

### 6. Data, 2UG 2015 HSC 4 MC

The data has been grouped into categories and because each category can be ranked, it is ordinal.

⇒ *A*

---

### 7. Data, 2UG 2017 HSC 4 MC

A systematic sample divides a population into equal sample sizes and then selects equally among them.

⇒ *C*

---

### 8. Data, 2UG 2018 HSC 3 MC

The number of brothers a person has is an exact whole number.

∴ Classification is numerical, discrete.

⇒ *C*

---

### 9. Data, 2UG 2007 HSC 17 MC

Total students surveyed

$$\begin{aligned} &= 10\% \times 1200 \\ &= 120 \end{aligned}$$

Students surveyed per year group

$$\begin{aligned} &= \frac{120}{6} \\ &= 20 \end{aligned}$$

A stratified sample would have sampled 10% of each year group.

$$\text{In Year 10, } 10\% \times 230 = 23$$

$$\text{In Year 11, } 10\% \times 150 = 15$$

$\therefore$  More students sampled in Year 10 and less in Year 11.

$\Rightarrow C$

### 10. Data, 2UG 2016 HSC 7 MC

Categorical and nominal data is qualitative and not ordered.

$\Rightarrow A$

♦♦ Mean mark 26%.

### 11. Data, 2UG 2009 HSC 24c

One example (of many):

The data might show that a large number of retirees live in a particular area. The government could decide to increase public transport in the area as the older retirees get, the more they rely on public transport.

### 12. Data, 2UG 2012 HSC 26d

(i) B, E, C, F, D, A

(ii) F

### 13. Data, 2UG 2006 HSC 23c

(i) A valid method would be using a stratified sample.

The number of students sampled in each year is proportional to the size of each year.

(ii)

Hours	Class Centre ( $x$ )	Frequency ( $f$ )	$fx$
0 - 4	2	69	138
5 - 9	7	72	504
10 - 14	12	38	456
15 - 19	17	21	357
Total			<b>1455</b>

**MARKER'S COMMENT:** This "routine" exercise of finding a mean from grouped data was incorrectly answered by most students! The best responses copied the table and inserted a class-centre column (see solution).

$$\begin{aligned} \text{Mean} &= \frac{\text{Sum of Scores}}{\text{Total scores}} \\ &= \frac{1455}{200} \\ &= 7.275 \text{ hours} \end{aligned}$$

#### 14. Data, 2UG 2011 HSC 25a

(i) Categorical

(ii) How many outgoing calls do you make per day?

(Ensure it can be answered with a numerical score.)

(iii) The method could be to work out how many students are in each year and ask 10% of the students in each year. (Note the sample of students in each year must be proportional to their percentage in the population).

◆◆ Mean mark 7%. Toughest mark to get in the 2011 exam!  
**COMMENT:** Know and be able to describe random, systematic and stratified sampling!

(iv) A census would involve all high school students in NSW.

◆◆ Mean mark 18%.  
**MARKER'S COMMENT:** A *specific* population needed (i.e. high school students).