

# Chapter 1 - Statistics

## Data Collection

### Chapter Overview

1. Populations and Samples
2. Sampling
3. Non-random Sampling
4. Types of Data
5. The Large Data Set

Topics	What students need to learn:	
	Content	Guidance
<b>1</b> <b>Statistical sampling</b>	<p><b>1.1</b></p> <p><b>Understand and use the terms 'population' and 'sample'.</b></p> <p><b>Use samples to make informal inferences about the population.</b></p> <p><b>Understand and use sampling techniques, including simple random sampling and opportunity sampling.</b></p> <p><b>Select or critique sampling techniques in the context of solving a statistical problem, including understanding that different samples can lead to different conclusions about the population.</b></p>	<p><b>Students will be expected to comment on the advantages and disadvantages associated with a census and a sample.</b></p> <p><b>Students will be expected to be familiar with: simple random sampling, stratified sampling, systematic sampling, quota sampling and opportunity (or convenience) sampling.</b></p>

## Populations and Samples

A **population** is –

A **sample** is –

A **census** is -

### Key words:

Individual units of a population are called \_\_\_\_\_. E.g. each individual person/item

Often sampling units of a population are individually named or numbered to form a list called a \_\_\_\_\_. E.g. the list of all people/items

### What are the advantages and disadvantages between taking a census and a sample?

	<b>Advantages</b>	<b>Disadvantages</b>
<b>Census</b>	<ul style="list-style-type: none"><li>• It should give a completely accurate result</li></ul>	<ul style="list-style-type: none"><li>• Time consuming and expensive</li><li>• Cannot be used when the testing process destroys the item</li><li>• Hard to process large quantity of data</li></ul>
<b>Sample</b>	<ul style="list-style-type: none"><li>• Less time consuming and expensive than a census</li><li>• Fewer people have to respond</li><li>• Less data to process than in a census</li></ul>	<ul style="list-style-type: none"><li>• The data may not be as accurate</li><li>• The sample may not be large enough to give information about small sub-groups of the population</li></ul>

**Example:** A supermarket wants to test a delivery of avocados for ripeness by cutting them in half.

- a. Suggest a reason why the supermarket should not test all the avocados in the delivery.

- b. The supermarket tests a sample of 5 avocados and finds that 4 of them are ripe. They estimate that 80% of the avocados in the deliver are ripe. Suggest one way that the supermarket could improve their estimate.

Sampling

Type	How to carry out	Advantages	Disadvantages
Simple Random Sampling	<p><b>What is it :</b></p> <p><b>Method:</b></p>		

Example. There are 64 girls and 56 boys in a school. Explain briefly how you could take a random sample of 15 pupils using a simple random sample. **(3)**

Type	How to carry out	Advantages	Disadvantages
Systematic Sampling	What is it :		

Example. A telephone directory contains 50 000 names. A researcher wishes to select a systematic sample of 100 names from the directory. Explain in detail how the researcher should obtain such a sample. **(2)**

Type	How to carry out	Advantages	Disadvantages
Stratified Sampling	What is it :		

Example. A school has 15 classes and a sixth form. In each class there are 30 students. In the sixth form there are 150 students. There are equal numbers of boys and girls in each class. There are equal numbers of boys and girls in the sixth form. The head teacher wishes to obtain the opinions of the students about school uniforms. Explain how the head teacher would take a stratified sample of size 40. **(7)**

## Non-random Sampling

Type	How to carry out	Advantages	Disadvantages
Quota Sampling			
Opportunity/ Convenience Sampling			

Example.

A lake contains 3 species of fish. There are estimated to be 1400 trout, 600 bass and 450 pike in the lake. A survey of the health of the fish in the lake is carried out and a sample of 30 fish is chosen.

- (a) Give a reason why stratified random sampling cannot be used. (1)
- (b) State an appropriate sampling method for the survey. (1)
- (c) Give one advantage and one disadvantage of this sampling method. (2)
- (d) Explain how this sampling method could be used to select the sample of 30 fish. You must show your working. (4)

## Types of Data

**Quantitative Data –**

**Qualitative Data –**

**Continuous Data –**

**Discrete Data –**

### **Key Vocabulary for Grouped Data (presented in a grouped frequency table)**

**Classes** - the groups in a table

**Class boundaries** - the maximum and minimum values in a class (beware of gaps between the classes)

**Midpoint** - the average of the class boundaries

**Class width** - the difference between the upper and lower class boundaries

Exercise 1D Page

### **Test Your Understanding**

#### **Name that sampling method**

Simple Random Sampling	Systematic Sampling	Stratified Sampling	Quota Sampling	Opportunity Sampling
------------------------	---------------------	---------------------	----------------	----------------------

1. "You wish to test lightbulbs produced by a factory in a daily batch."
2. "You wish to survey consumer opinion on your new drink *FizzGuzz* released in the UK."
3. "You wish to determine students' favourite TV programmes in your school, that is fairly representative of each year group."

## The Large Data Set

All A Level exam boards are obligated to provide a 'large data set'. Data in exam questions will often be from this set, and you are encouraged to explore this data (which is publicly available) in Microsoft Excel.

**It is important to note that you are expected to be familiar with this data set before you go into your exam, including some basic geographic knowledge!**

Date	Daily Max Temp (09-00-0900 C)	Daily Total Rainfall (0900-0900) (mm)	Daily Total Sunshine (0000-2400) (hrs)	Daily Mean Windspeed (0000-2400) (kn)	Daily Mean Windspeed (0000-2400) (Beaufort conversion)	Daily Maximum Gust (0000-2400) (kn)
01/06/1987	15.1	0.6	4.5	7	Light	19
02/06/1987	12.5	4.7	0	7	Light	22
03/06/1987	13.8	tr	5.6	11	Moderate	25
04/06/1987	15.5	5.3	7.8	7	Light	17
05/06/1987	13.1	19	0.5	10	Light	33
06/06/1987	13.8	0	8.9	19	Fresh	46
07/06/1987	13.2	tr	3.8	11	Moderate	27
08/06/1987	12.9	1	1.7	9	Light	19
09/06/1987	11.2	tr	5.4	6	Light	19
10/06/1987	9.2	1.3	9.7	4	Light	n/a
11/06/1987	12.6	0	12.5	6	Light	18
12/06/1987	10.4	0	11.9	5	Light	n/a
13/06/1987	9.6	0	8.6	5	Light	15
14/06/1987	10.2	0	13.1	5	Light	18
15/06/1987	9.2	3.7	7.1	4	Light	25
16/06/1987	10.4	5.6	8.3	6	Light	25
17/06/1987	12.8	0.1	5.3	10	Light	27
18/06/1987	13.0	7.4	3.2	9	Light	24
19/06/1987	14.0	tr	0.4	12	Moderate	33
20/06/1987	12.6	0	7.7	6	Light	17

Example.

- (a) Describe the type of data represented by daily total rainfall.

Alison is investigating daily maximum gust. She wants to select a sample of size 5 from the first 20 days in Hurn in June 1987. She uses the first two digits of the date as a sampling frame and generates five random numbers between 1 and 20.

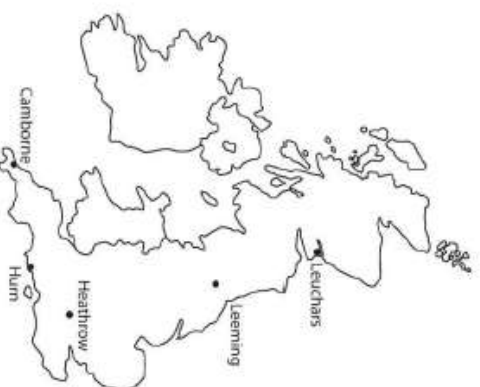
- (b) State the type of sample selected by Alison.
- (c) Explain why Alison's process might not generate a sample of size 5.



## The Large Data Set

### Locations

5 UK weather stations



### Time Periods

May – October 1987 (6 months)  
May – October 2015 (6 months)

### Seasons

May/June are the end of spring  
July-Sept is summer  
October is autumn

Perth (Australia) is in the southern hemisphere, so July-Sept is winter

### UK Great Storm

The night of 15-16<sup>th</sup> October 1987  
Gusts up to 100 knots recorded

### Florida hurricanes

12 October 1987 Hurricane Floyd  
1-2 October 2015 Hurricane Joaquin

3 Overseas



### Variables Recorded

**Daily Maximum Temperature**  
°C

**Daily Total Rainfall**  
mm

**Daily Total Sunshine**  
hours

**Daily Maximum Relative Humidity**  
%; mist and fog if > 95%

**Daily Mean Windspeed;**  
**Daily Maximum Gust**  
knots (1 kn = 1.15mph)  
and Beaufort scale

**Daily Mean Wind Direction;**  
**Daily Maximum Gust Direction;**  
bearing (°)  
and cardinal direction

**Cloud Cover**  
oktas (eights): 0 – 8

**Visibility**  
Dm (decametres)  
1 Dm = 10m

**Pressure**  
hPa (hectopascal)

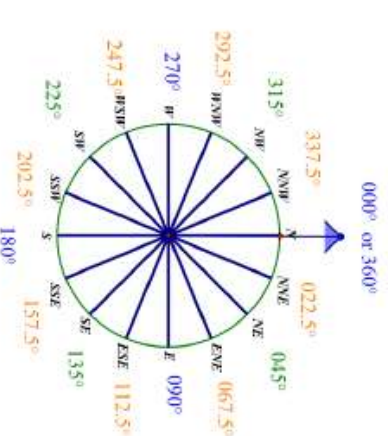
n/a  
reading not available

tr (trace)  
rainfall < 0.05mm

### Beaufort Scale

Discrete, scale of 13 values:  
0 (calm, < 1kn)  
12 (hurricane, 64kn+)

### Cardinal Directions



### Oktas

Eighths of the sky covered by cloud  
Discrete, scale of 9 values:  
0 (clear sky)  
8 (completely overcast)

Sources  
Maps: Pearson  
Compass: mathsmutt.co.uk