

# Chapter 1 - Data Collection

1.1 - Populations and Samples - Pg. 2

1.2, 1.3 - Sampling & Non-random Sampling - Pg. 3 - 5

1.4 - Types of Data - Pg. 6

1.5 - The large data set - Pg. 7 - 9

---

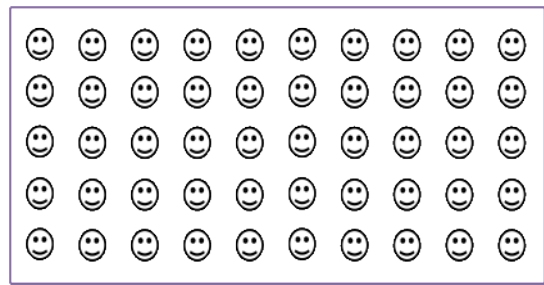
Personal notes:



# 1.1 - Populations and Samples

## Key Terms

- A population is
- A sample is
- A sampling unit is
- A sampling frame



## Notes

- We could collect data from either a sample, or from the entire population. Data collected from the entire population is known as a \_\_\_\_\_.

## Example

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

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

## 1.2, 1.3 - Sampling & Non-random Sampling

- There are two types of sampling: Random and Non-Random.
- Each type of sampling has a few methods.



## 1.2 - Random Sampling

A random sampling is such that each thing in the sampling frame to have an equal chance of being chosen, in order to **avoid bias**.

### Simple Random Sampling (Random)

Edexcel S3 June 2004 Q1a

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)

### Systematic Sampling (Random)

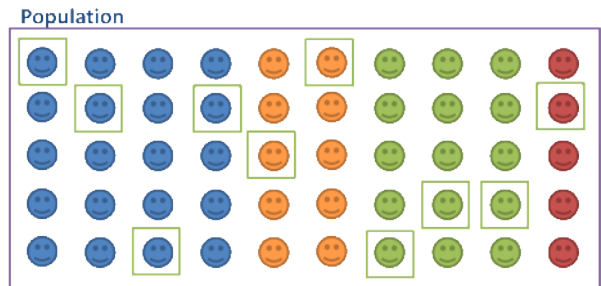
Edexcel S3 June 2009 Q1a

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)

### Stratified Sampling (Random)

For example, we want to sample 20% of the population (shown as the diagram). If the population were divided into distinct group (4 colours = 4 groups, say age), known as "**strata**".

We could randomly sample 20% from each group, ensuring each group is equally represented.



Edexcel S3 Jan 2006 Q1

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)



## 1.3 - Non-Random Sampling

Consider this scenario:

You are going to conduct a survey on whether ability in mathematics correlates the person's innate art talent. We need to choose some mathematicians to assess.



Why would random sampling be problematic?

Solution: For this scenario, \_\_\_\_\_ should be deployed.

1. As with stratified sampling, divide population into groups according to characteristic of interest, then determine size of each group in sample to reflect proportions within the population.
2. But instead of random sampling within each group, we actively choose people within each group via suitable means (e.g. advertising), until the 'quota' for each group is filled.

Another type of non-random sampling is \_\_\_\_\_ (aka convenience sampling).

- It consists of taking the sample from people who are available at the time the study is carried out and who fit the criteria you are looking for.
- E.g. Survey about relationship between age and likeliness to use re-usable bags  
-> Could collect data from the first 20 people coming out from a supermarket

### Example

#### Edexcel S3 June 2010 Q2

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)



## 1.4 - Types of Data

### Notes

- Data can be presented in a \_\_\_\_\_ for conciseness.

Shoe size, $w$	Frequency
$2 \leq w < 5$	3
$6 \leq w < 11$	4

$$\begin{array}{c} \nearrow 6 \leq w < 11 \searrow \\ \text{---} \end{array}$$



# 1.5 - The large data set

Starter: Suggest a suitable sampling method for each of the following scenarios.

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

“You wish to test lightbulbs produced by a factory in a daily batch.”

“You wish to survey students’ opinion on Mr. Fan’s new drawing *Geometry*.”

“You wish to determine Londoners’ favourite TV programmes.”

## Basic Information

- 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**

**Introduction**  
Pearson have provided this large data set, which will support the assessment of Statistics in the A Level Mathematics Paper 3 and AS Mathematics Paper 2. Students are expected to become familiar with the data set in advance of the final assessment.

To support the use of the large data set in the teaching of the statistics content, tasks include:

- selecting a sample
- cleaning the data
- creating a graph from the data
- calculating summary statistics such as mean, standard deviation
- calculating regression equations and correlation coefficients where applicable
- hypothesis testing.

It is expected that students should use technology such as spreadsheets or other statistical packages to explore the data.

See the specifications A Level Mathematics (IPMA) and AS Mathematics (IPMA2) for further information.

**Data set source**  
The data set consists of weather data samples provided by the Met Office for five UK weather stations and three overseas weather stations in the time periods May to October 1987 and May to October 2010. The weather stations are labelled on the maps shown:  
- in the UK - Cambridge, Heathrow, Hull, Leamington and Leuchars  
- overseas - Beijing, Jask and Perth.  
Further information about our data source can be accessed at <http://www.metoffice.gov.uk/>

**Dataset variables and explanatory notes**  
The Met Office provides data for a number of different weather variables. Our data set includes data for eleven variables recorded across the weather stations during the set periods of time.

**Daily Mean Temperature**  
All temperatures are recorded by thermometers in a hollowed screen 1.25 metres above short grass, except at some Weather Centre's and Climate Data Logger stations, where observations are made from a sheltered and open area.

Information | Cambridge May-Oct 1987 | Heathrow May-Oct 1987 | Hull May-Oct 1987 | Leamington May-Oct 1987 | Leuchars May-Oct 1987 | Cambridge M

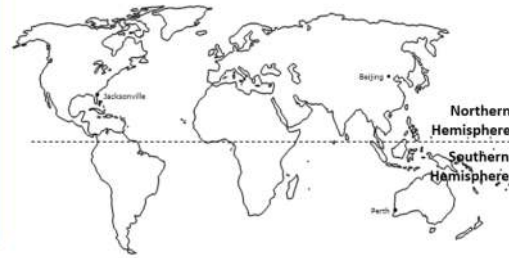
Edexcel's data set concerns **weather data** from a number of weather stations. Let's explore what you might be expected to know...



# 1.5 - The large data set

## What you'll need to be familiar with

- Names and rough locations of the 5 UK weather stations, as well as the 3 international weather stations.



The data was recorded for:  
 - May to Oct 1987  
 - May to Oct 2015

- Variables involved and their respective units.

**Total rainfall**  
(in mm)

tr/trace means less than 0.05mm

**Mean Windspeed**

kn/knot is "nautical mile per hour". 1kn = 1.15 mph  
 Windspeed also given on Beaufort Scale:

0 = Calm	< 1kn
1-3 = Light	1-10kn
4 = Moderate	11-16kn
5 = Fresh	17-21kn

**Mean Visibility**

How far (in metres) can be seen into the horizon during daylight hours.

CAMBORNE © Crown Copyright Met Office 2015													Wind Direction	
NR = 1627E 4067N Altitude = 87 metres Latitude = 50:22N Longitude = 05:33W														
Date	Daily Mean Temperature (0900-0900) [°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)	Daily Maximum Relative Humidity %	Daily Mean Total Cloud (oktas)	Daily Mean Visibility (Dm)	Daily Mean Pressure (hPa)	Daily Mean Wind Direction (e)	Cardinal Direction	Daily Max Gust Corresponding Direction (e)	Cardinal Direction
01/05/1987	10.7	3.1	n/a	n/a	n/a	n/a	100	7	2000	1018	360	N	30	NNE
02/05/1987	8.9	0.1	n/a	n/a	n/a	n/a	91	3	2200	1020	320	NW	330	NNW
03/05/1987	8.1	0	n/a	n/a	n/a	n/a	77	5	3900	1029	350	N	350	N
04/05/1987	8.2	0	n/a	n/a	n/a	n/a	83	5	4100	1036	350	N	350	N
05/05/1987	9.8	0	n/a	n/a	n/a	n/a	86	5	2700	1036	10	N	10	N
06/05/1987	9.3	0	n/a	n/a	n/a	n/a	100	1	1000	1033	290	W	260	W
07/05/1987	10.9	0	n/a	n/a	n/a	n/a	100	3	600	1021	350	NW	310	NW
08/05/1987	10.5	tr	n/a	n/a	n/a	n/a	89	1	2400	1025	110	WNW	340	WNW
09/05/1987	10.9	0	n/a	n/a	n/a	n/a	95	3	900	1017	360	N	350	N
10/05/1987	9.9	0	n/a	n/a	n/a	n/a	79	4	4100	1018	10	N	10	N
11/05/1987	8.8	6	n/a	n/a	n/a	n/a	95	7	2500	1017	270	W	260	W
12/05/1987	10.2	tr	n/a	n/a	n/a	n/a	97	5	2400	1009	340	NW	310	NW
01/05/1988	11.6	2.2	19	16	Moderate	39	77	4	4600	1016	340	NNW	340	NNW
02/05/1988	11.6	tr	19	16	Moderate	39	95	7	3100	1008	290	WNW	270	W
03/05/1988	11.6	tr	12.3	13	Moderate	23	77	4	4500	1012	10	N	10	N
04/05/1988	11.6	tr	11.6	6	Light	14	92	4	3700	1015	290	NNW	290	WNW

**Mean temperature**  
(in °C)

Textbook claims this is max temp for UK, but it is mean temp for all locations.

**Total sunshine**  
(nearest 1/10 of an hour)

**Maximum Gust**  
(in kn) is highest instantaneous wind speed.

**Humidity**  
is the % of air saturation with water vapour. 100% is the maximum % water content air can contain.

**Mean Cloud Cover**  
Oktas means the number of 1/8ths of the sky covered.

- A vague idea of the range of values for each location.  
 Example: Mean wind speed in UK across full period was roughly 9 nm. But 4 nm in Beijing (i.e. lower), 5 in Jacksonville (again lower), 8 in Perth (similar to UK).
- A vague idea of the range of values of each variable.

Variable	Typical value(s)
Gust (UK only)	8 – 52 nm
Rainfall	0 – 60 mm in UK, but more extreme maximums elsewhere (e.g. 102mm in Perth)
Pressure	988 – 1038 hPa
Wind Speed on Beaufort scale	Max is 'fresh' (5). Most Light or Moderate.
Sunshine (UK only)	0 – 16 hrs
Cloud Cover	0 – 8 oktas (i.e. full spread)

**\*\*Note:**

- If you need to do calculations on the large data set in your exam, the relevant extract from the data set will be provided.
- You will need to recall trends from within the data set, or identify a location based on the given data.

## 1.5 - The large data set

### Example

Answer the following questions using the extract provided from the large data set

- 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

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

HURN						
© Crown Copyright Met Office 1987						
Date	Daily mean temperature (°C)	Daily total rainfall (mm)	Daily total sunshine (hrs)	Daily mean windspeed (kn)	Daily mean windspeed (Beaufort conversion)	Daily maximum gust (kn)
01/6/1987	15.1	0.6	4.5	7	Light	19
02/6/1987	12.5	4.7	0	7	Light	22
03/6/1987	13.8	tr	5.6	11	Moderate	25
04/6/1987	15.5	5.3	7.8	7	Light	17
05/6/1987	13.1	19.0	0.5	10	Light	33
06/6/1987	13.8	0	8.9	19	Fresh	46
07/6/1987	13.2	tr	3.8	11	Moderate	27
08/6/1987	12.9	1	1.7	9	Light	19
09/6/1987	11.2	tr	5.4	6	Light	19
10/6/1987	9.2	1.3	9.7	4	Light	n/a
11/6/1987	12.6	0	12.5	6	Light	18
12/6/1987	10.4	0	11.9	5	Light	n/a
13/6/1987	9.6	0	8.6	5	Light	15
14/6/1987	10.2	0	13.1	5	Light	18
15/6/1987	9.2	3.7	7.1	4	Light	25
16/6/1987	10.4	5.6	8.3	6	Light	25
17/6/1987	12.8	0.1	5.3	10	Light	27
18/6/1987	13.0	7.4	3.2	9	Light	24
19/6/1987	14.0	tr	0.4	12	Moderate	33
20/6/1987	12.6	0	7.7	6	Light	17

### Example

Calculate (using the same extract):

- The mean daily maximum temperature for the first five days of June in Hurn in 1987.
- The median daily total rainfall for the week of 14<sup>th</sup> June to 20<sup>th</sup> June inclusive.
- The median daily total rainfall for the same week in Perth was 19.00mm. Karl states that more southerly countries experience higher rainfall during June. State with a reason whether your answer to part (b) supports this statement.

