

### Topics from GCE AS and A Level Mathematics covered in Sections 4:

- Understand and use the terms 'population' and 'sample'.
- Understand and use sampling techniques including random sampling, systematic sampling and opportunity sampling.
- 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.

### The problem

A researcher is interested in the training regime of swimmers who took part in the 2012 Summer Olympics. To investigate this he would, ideally, like to collect extra information for all the Olympic swimmers who took part in these Olympic Games.

As it would be too time-consuming to collect this extra information for all 907 Olympic swimmers a sample will be selected. He has asked you to select samples of 50 Olympic swimmers from the 907 who took part in the 2012 Summer Olympics using three different sampling methods:

- simple random sampling
- systematic sampling
- opportunity sampling.

He also asks for a report on the advantages and disadvantages of each sampling method. For each of the three sampling methods the report should include:

- a definition of each method
- an efficient way of selecting a sample
- the advantages and disadvantages of each method in context.

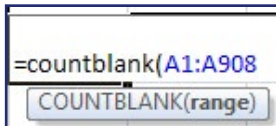
### Data collection

The data are provided in the Excel worksheet '**2012OlympicSwimmers.xlsx**'. Open the workbook and select the worksheet **Swimmers**. This contains data for the Olympian swimmers only.

### Process

Check if there are any missing values in the Name column.

**To use the COUNTBLANK function to see if any Olympian swimmers' names are missing**



- In the worksheet **Swimmers**, select **R2** type **=count**.

A list of COUNT functions will appear.

- Double click on **COUNTBLANK**.

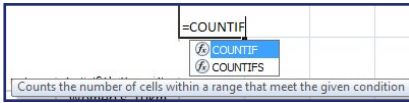
This counts the number of empty cells in a specific range.

- Select **A1** to **A908** (use the cursor to select these cells or type in the cell references).
- **Enter**
- **Do not select the whole of column A** as there are thousands of empty cells in the whole column.
- **Save your work**

### 1. Are there any names missing?

*There are no missing names.*

### To use the COUNTIF function to count the number of males and females

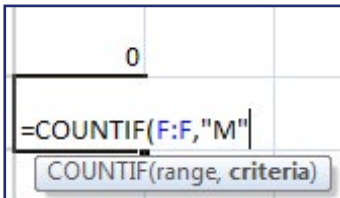


- Select **Q3** type **Male**
- Select **R3** type **=count**

A list of COUNT functions will appear.

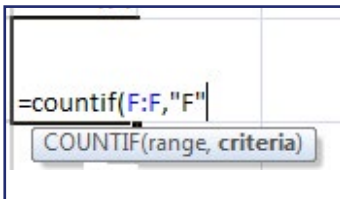
- Double click on **COUNTIF**

This function counts the number cells within a range that meet a given condition.



Count the number of males:

- Select column **F**
- Type , **"M"** (a comma followed by "M"), **ENTER**.
- The function will count the number of Ms in the column.
- The quotation marks are needed as M is text and not **number**.



- Select **Q4** type **Female**
- Select **R4** and use the **COUNTIF** function to find the number of females
- Convert these frequencies to percentages in **S3** and **S4**.
- **Save your work**

### 2. Complete the table.

2012 Olympic swimmers		
Sex	Frequency	Percentage
Male	474	52
Female	433	48

### Some definitions

#### Target population

The target population is all the members of the population you would ideally like to get the information from.

#### 3. State the target population for this investigation?

*Olympic swimmers who took part in the 2012 Summer Olympics.*

#### Sampling frame

A sampling frame is a list of all the target population.

#### 4. Is there a sampling frame for this investigation?

*For this investigation there is a sampling frame. But in many cases this is not available.*

#### Bias

Bias is a systematic error. For example, if your sampling method is to select every fourth Olympian but for some reason the Olympians are listed so every fourth one is male, then this is an error in the system used to select the sample and the sample will be biased. The sample should be as representative of the target population as possible so bias needs to be eliminated.

#### 5. Give another example of how an error in a sampling system could arise.

*If your sampling method is to select every 10th Olympian and these were all, for example, medal winners, from the same country. Or the researcher selected the Olympic swimmers he had heard of.*

#### Sampling error

There is always a sampling error, as a sample is always part of the population being investigated. A sample is used to make inferences and generalisations about the population. It is important to understand that different samples will give different outcomes.

Variation between samples cannot be eliminated, but researchers can aim to eliminate bias.

#### 6. Will the samples you select have sampling errors?

*Yes all samples have sampling errors.*

### Sampling Methods

#### Simple random sampling

Simple random sampling is where every member of the population has the same probability of being selected. This method relies on having a sampling frame. It can be time consuming and expensive to carry out.

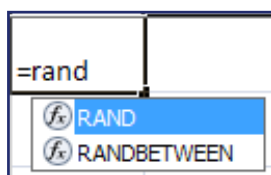
One way to select a simple random sample of the 2012 Olympic swimmers is to write the name of each Olympian on a separate piece of identical paper, fold these identically, put them in a hat and randomly select 50. Of course, this is impractical and very time consuming.

Another way is to number each Olympic swimmer from 001 to 907 and then refer to a random numbers table or generate (using software or a calculator) a three digit random number. Then match this three digit random number to the number allocated to the Olympic swimmer. This is the first member of your sample.

Repeat this until 50 Olympians have been selected. Discard random numbers 000 and 908–999 and select the sample without repetition.

This method is also quite time consuming and a quicker way to select a simple random sample is by using the RAND function in Excel.

### To select a random sample using the RAND function in Excel

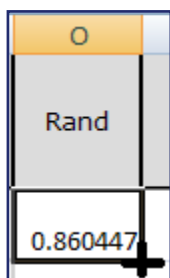


In the worksheet **Swimmers**:

- select **O1** and type the header **Rand**
- select **O2** and type **=rand()** and **Enter**.

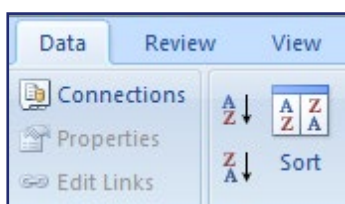
The RAND function returns a random number greater than or equal to 0 and less than 1, evenly distributed.

Note: the random number will change every time the key F9 is pressed or when a recalculation takes place in the worksheet.



Copy the rand() function in every cell in the **O** column where there is an entry in the **N** column:

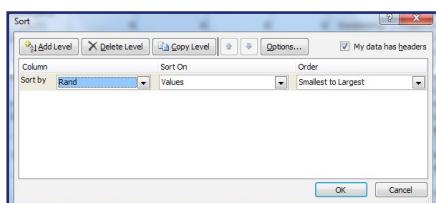
- Select **O2** and hover the cursor over the bottom corner of the cell until a small black cross appears and double click (or drag the cursor down to **O908**).



Sort the data according to column **O**:

- Select columns **A** to **O**
- Select the **Data** tab
- Select **Sort**
- Tick **My data has headers**
- Sort by **Rand**
- **OK**.

You may notice the random numbers in the **Rand** column do not look in order. This is because new random numbers were generated by using the **Sort** function.



- Select the first 50 Olympians (up to row 51 including headers).
- Select **A1** to **N51** right click **Copy**.
- Select the worksheet **Simple random sample**.
- Select **A1** and **Paste**.

Remember every student will have a different sample.

- **Save your work**

### 7. Does it matter if you sort the rand column from smallest to largest or largest to smallest?

*No as the sample is a simple random sample.*

### Systematic sampling

Systematic sampling is where a starting point is chosen at random and objects are then selected at fixed periodic intervals.

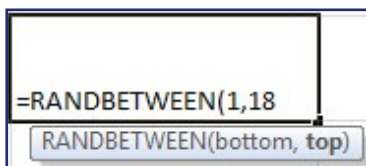
From a population of size 907 a sample size of 50 will be selected.

Divide 907 by 50 which is approximately equal to 18.

Start by selecting an Olympic swimmer at a random starting point, followed by every 18th Olympic swimmer from then on.

### To select a systematic sample in Excel

*There are many ways of selecting a systematic sample in Excel. The method detailed below is a simple procedure to help students understand the selection process.*



- In the worksheet **Swimmers**, select **Q1** and type **=rand**
- Double click on **RANDBETWEEN**

This function will generate a random number between two specified numbers (inclusive).

- Type **1,18** and **Enter**

This is the starting point.

*For example:*

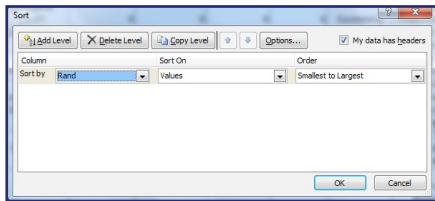
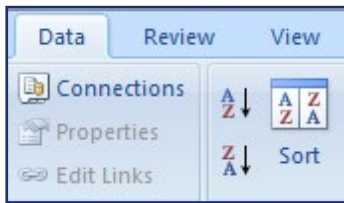
For the random number 11 generated from the function RANDBETWEEN, the starting point is the 11<sup>th</sup> Olympic swimmer in the sampling frame.

*(The order of the data may be different for each student as they have just sorted the data following use of the RAND function to select a simple random sample.)*

To select the 11<sup>th</sup> Olympian:

- Select **P12**, type **1** and **Enter**
- Count down 18 Olympians (or cells) to **P30** and type **2** into this cell
- Select **P13** to **P30**, hover the cursor over the bottom right hand corner until a black cross appears and **double click**.

This has copied the pattern in column P in all the cells where there are data in column N. So, in column P, there are consecutive numbers in cells that are 18 cells apart.



Sort the data according to column P:

- Select columns **A to P**
- Select the **Data** tab
- Select **Sort**
- Tick **My data has headers**
- Sort by **Column P**
- Order **Smallest to Largest**
- **OK.**

- Select the first 50 Olympians (up to row 51 including headers).
- Select **A1 to N51** right click **Copy**.
- Select the worksheet **Systematic sample**.
- Select **A1** and **Paste**.

Remember every student will have a different sample.

- **Save your work**

### Opportunity sampling

Opportunity sampling involves taking the sample from participants who are available and willing to take part at the time the investigation is being carried out and who fit the necessary criteria.

This method is easy and inexpensive to carry out, but may not be representative of the target population as it could be subject to bias.

The extra information the researcher needs to collect has been found on a sports website. The names of the Olympic swimmers for whom this extra information can be collected have been pasted, along with the other variables, in the worksheet **Opportunity sample**.



### Process

Inspect the samples to see if they are representative of the target population.

### 8. Complete the table below.

Find the number of medal winners in the sample by either:

- sorting the data according to the variable '**Total number of medals**' and counting the number of Olympic swimmer medal holders;

or

- using the COUNTIF function i.e. =COUNTIF(L:L,"> 0")

	Sample size	Frequency		Percentage		Number of medal winners
		Male	Female	Male	Female	
Population	907	474	433	52	48	97
Simple random sample						
Systematic sample						
Opportunity sample	50	24	26	48	52	50

### Report

#### 9. Write a report.

The report should include:

- A definition of each sampling method
- An efficient way of selecting a sample using each sampling method
- The advantages and disadvantages of each sampling method in this context
- Students must refer to the samples selected as evidence to back up their advice.

Sampling Method	Advantages	Disadvantages
<i>Simple random</i>	<p><i>Representative if the sampling frame is available.</i></p> <p><i>Can be quick and easy to use.</i></p>	<p><i>Not possible without a sampling frame.</i></p> <p><i>Potentially time-consuming and expensive.</i></p>
<i>Systematic</i>	<p><i>As for simple random sampling</i></p>	<p><i>Not possible without a sampling frame.</i></p> <p><i>If the sampling technique coincides with a periodic trait in the population, the sampling technique will no longer be random or representative. This would introduce bias.</i></p>
<i>Opportunity</i>	<p><i>Quick and easy. Convenient.</i></p>	<p><i>May not be representative of the target population and could be biased</i></p>

*Discuss the variation between samples.*

*Make the point that each student has selected a different sample.*

*Neither simple random sampling nor systematic sampling is the best in this case and the opportunity sample is biased.*