Topics from GCE AS and A Level Mathematics covered

* Understand and use the terms ‘population’ and ‘sample’.
* Advantages and disadvantages associated with a census and a sample.
* Understand and use sampling techniques, including simple random sampling and opportunity sampling. Students will be expected to be familiar with: simple random sampling, stratified sampling, systematic sampling, quota sampling and opportunity (or convenience) 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.

**Investigation 3**

A couple going on holiday, flying from Heathrow, wish to avoid the expensive times of the year. They decide to select a sample of 30 dates during May to October to check out the prices of return flights to Beijing.

**The data**

The data are provided in the Excel spreadsheet **Dates.xlsx** Open the workbook and select the worksheet **Dates.** Thiscontains dates and days of the week between May and October 2018.

1. State the target population for this investigation?

The target population is all the members of the population you would ideally like to get the information from. In this case it is all the dates from May to October 2018 (inclusive).

1. Is there a sampling frame for this investigation?

A sampling frame is a list of all the target population. For this investigation there is a sampling frame, the list of all dates from May to October 2018. But in many cases this is not available.

1. What is bias?

Bias is a systematic error. 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.

1. Will the samples you select have sampling errors?

There is always a sampling error as a sample is always part of the population being investigated. A sample is being used to know something about the population.

It is important to understand that different samples will give a slightly different picture. Variation between samples cannot be eliminated but researchers can aim to eliminate bias.

**Simple random sampling**

A simple random sample is when every possible sample of given size 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 dates is to write the date on a separate piece of identical paper, fold these identically, put them in a hat and randomly select 30. Of course, this is impractical and very time consuming.

Another way is to number each date from 001 to 184 (or 000 to 183) and then refer to a random numbers table or generate (using software or a calculator) a three digit random number. Match this three digit random number to the number allocated to the date. This is the first member of your sample.

Repeat this until 30 dates have been selected. Discard random numbers 000 and 185 – 999 (or 183 to 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.

U*sing the RAND function in Excel to select a simple random sample*

*In the worksheet* ***SimpleRandom***

*Select* ***D2*** *type* ***=RAND()*** *and* ***enter***

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

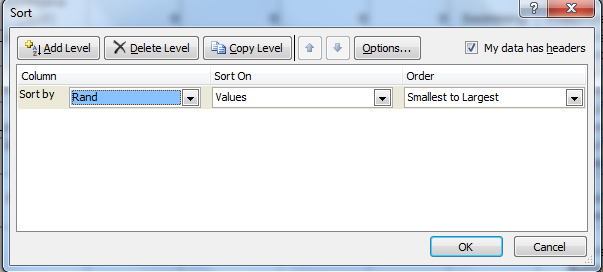
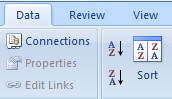
*Note: the random number will change every time the key F9 is pressed or when a recalculation takes place in the spreadsheet.*

*Copy the RAND() function in every cell in the* ***D*** *column where there is an entry in the* ***C*** *column*

*Select* ***D2*** *and hover the cursor over the bottom right hand corner of the cell until a small black cross appears and double click (or drag the cursor down to* ***D185****)*

*Sort the data according to column* ***D***

*Select columns* ***A*** *to* ***D***

*Select the* ***Data*** *tab*

*Select* ***Sort***

*Tick* ***My data has headers***

***Sort by******Rand***

***OK***

*The random numbers in the* ***Rand*** *column do not look in order. This is because new random numbers were generated, after the sort, by using the* ***Sort*** *function.*

*Select the first 30 dates (up to row 31 including headers) and copy into columns* ***F:H***

*Select* ***A1*** *to* ***C31*** *right click* ***Copy***

*Select* ***F1*** *right click* ***Paste***

*Inspect the frequency tables for the frequencies of each day, month and time period in the sample. Remember every student will have a different sample.*

***Save your work***

1. Within the **Sort** window 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 the numbers are in a random order either way.

**Systematic sampling**

A systematic sample is when you choose a starting point at random then systematically select objects at a certain number apart.

1. Explain why, in this case, it would not be a good idea to select a systematic sample by choosing a random starting point and then selecting every 7th member of the population.

The sample members would all be the same day of the week. Holiday prices may vary according to the day of the week you start the holiday. This would result in the sample being biased.

From a population of size 184 a sample size of 30 will be selected.

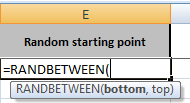
Divide 184 by 30 which is approximately equal to 6.

Start by selecting the date at the random starting point followed by every 6th date from then on. Of course, the starting points will vary between students.

*To select a systematic sample in Excel*

There are many ways of selecting a systematic sample in Excel. The way explained here does it in a simple way to help students understand the selection process.

*To select the random starting point*

*In the worksheet* ***Systematic***

*Select* ***E2*** *type* ***=rand***

*Double click**on* ***RANDBETWEEN***

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

*Type* ***1,6 Enter***

*Example*

*For the random number 5 generated from the function RANDBETWEEN. The starting point is the 5th date in the sampling frame.*

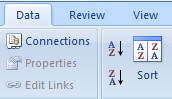
*To select the 5th date and the 6th from then on*

*Select* ***D6*** *type* ***1******Enter*** *(The 5th value is in D6.)*

*Count down 6 dates (or cells) from* ***D6*** *to* ***D12*** *type* ***2*** *into this cell*

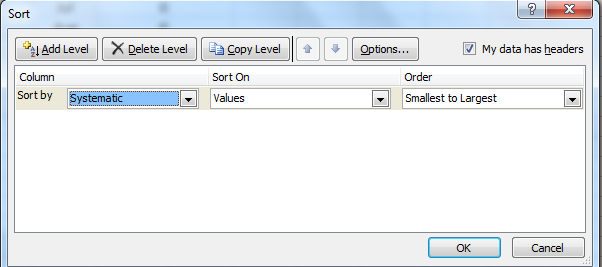
*Select* ***D7*** *to* ***D12 (****do not include D6) hover the cursor over the bottom right hand corner until a black cross appears and* ***double click***

*This has copied the pattern in column D in all the cells where there are data in the cells in column C. So in column D there are consecutive numbers in cells that are 6 dates (or cells) apart.*

*Sort the data according to column D*

*Select columns* ***A*** *to* ***D***

*Select the tab* ***Data*** *then* ***Sort***

*Tick* ***My data has headers***

***Sort by Column D (Systematic)***

*Order* ***Smallest to Largest***

***OK***

*Select the first 30 dates (up to row 31 including headers) and copy into columns* ***F:H***

*Select* ***A1*** *to* ***C31*** *and**right click* ***Copy***

*Select* ***G1*** *and right click* ***Paste***

*Inspect the frequency tables for the frequencies of each day, month and time period in the sample. Remember every student will have a different sample.*

***Save your work***

**Stratified sampling**

A stratified sample is when the population is split into distinguishable groups which are quite different from each other and which together cover the whole population. These groups are called strata. Within each group, or stratum, a probability sample is selected. The frequencies for each group in the sample are often proportional to the frequencies for each group in the population – this is proportional stratified sampling.

The couple going on holiday wish to split the 6 months into two strata, May to August being one and September to October the other.

There are 184 dates in the population. There are 123 dates during May to August so the proportion of these in the sample is 123 divided by 184. Multiply this by 30 to give the frequency of dates in May to August in the stratified sample of size 30.

|  |  |  |  |
| --- | --- | --- | --- |
| Stratum | Number of days | Calculation | Stratified sample frequency |
| May - August | 123 |  | 20 |
| September - October | 61 |  | 10 |

*Select a simple random sample of size 20 from the dates between May and August inclusive.*

*In the worksheet* ***Stratified*** *the data have been split into May to August and September to October.*

*Select* ***D2*** *type =****RAND() Enter***

*Copy this down to* ***D124***

*Sort columns* ***A:D*** according to column ***D (Rand)***

*Copy and paste columns* ***A1:C21*** *into* ***K1:M21***

*Select a simple random sample of size 10 from the dates between September and October inclusive.*

*Enter the RAND function in* ***I2.***

*Copy this down to* ***I62***

*Sort columns* ***F:I*** according to column ***I (Rand)***

*Copy and paste columns* ***F2:H11*** *into* ***K22:M31***

*This is the stratified sample.*

*Inspect the frequency tables for the frequencies of each day, month and time period in the sample. Remember every student will have a different sample.*

***Save your work***

**Quota sampling**

A quota sample is when the population is split into groups or strata as for stratified sampling. Then a judgement is used to select the members from each group

This is often used by companies who employ people to carry out opinion surveys. The size of the quota is worked out in the same way as for stratified sampling but the choice of who is sampled is then left up to the interviewer and so this method is definitely not random. A quota sample is not included here.

**Opportunity sampling**

Opportunity sampling consists of taking the sample from members who are accessible. 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. An opportunity sample is not included here.

**Process**

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

1. Complete the tables below for your samples.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Frequency** | | | | | | |
| **Days** | **Mon** | **Tue** | **Wed** | **Thu** | **Fri** | **Sat** | **Sun** |
| Population |  |  |  |  |  |  |  |
| Simple random sample |  |  |  |  |  |  |  |
| Systematic sample |  |  |  |  |  |  |  |
| Stratified sample |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Frequency** | | | | | | | |
| **Months** | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** | **May to Aug** | **Oct to Sep** |
| Population |  |  |  |  |  |  |  |  |
| Simple random sample |  |  |  |  |  |  |  |  |
| Systematic sample |  |  |  |  |  |  |  |  |
| Stratified sample |  |  |  |  |  |  |  |  |

**Report**

1. 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.
* Refer to the samples selected as evidence to back up the advice.

|  |  |  |
| --- | --- | --- |
| **Sampling Method** | **Advantages** | **Disadvantages** |
| Simple random | Representative if the sampling frame is available.  Can be quick and easy to use. | Not possible without a sampling frame.  Potentially time consuming, disruptive and expensive.  There may be missing values in the sample. |
| Systematic | As for simple random sampling. | Not possible without a sampling frame.  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.  There may be missing values in the sample. |
| Stratified | Representative if the sampling frame is available.  The frequencies for each group in the sample can be proportional to the frequencies for each group in the population | Not possible without a sampling frame.  Strata must be carefully defined.  Sometimes difficult to split the population into naturally occurring groups.  There may be missing values in the sample. |
| Quota | As for stratified sampling.  Does not need a sampling frame. | Non-random.  May not be representative of the target population and could be biased. |
| Opportunity | Quick and easy. Convenient. | Non-random.  May not be representative of the target population and could be biased. |

Discuss the variation between samples.

Make the point that each student has selected different samples and therefore may have different conclusions. Their conclusions must be backed up with evidence or reasons.