

SAMPLING

When working with data we have to ensure that we collect relevant information. We may need to collect data from a large population but it may not be possible to gather this information from everyone. This could be due to time, money or other constraints. Therefore, it is useful to use a sample of the population.

Check first that you:

- understand the difference between qualitative data (words) and quantitative data (numbers)
- can identify what information is needed to test a hypothesis
- can design and criticise questions for a questionnaire
- understand fairness and bias.

Sampling techniques In order to ensure that we get a fair and non-biased sample, we use one of the following sampling techniques:

Simple random sampling

- This is when each member of the population has the same chance of being selected for the sample.
- The sample may be chosen by drawing names from a hat.
- It could also be chosen by giving all individuals in the population a number and then using a calculator, computer or random number tables to generate numbers for individuals to be chosen.

Systematic random sampling

- This is very similar method to random sampling but the population would first be ordered according to specific criteria such as listing names of people in the population in alphabetical order.
- The sample would be drawn by selecting every n^{th} person. For example every 10th person in the list.

Stratified random sampling

- This method of sampling is used when each member of the population can be distributed into a certain group such as gender or age.
- A random sample is taken from every group but the number of members selected from each of the groups should be in proportion to the size of the group within the whole population.
- The number selected from each group is given by:

$$\text{No. selected from group} = \frac{\text{size of group}}{\text{size of population}} \times \text{size of sample}$$

Remember that the population is the whole of the data but a sample is just a part of this data.

Remember to ensure that the sample size is large! This is important to ensure that it is representative of the whole population we are gathering information on.

Simple random sampling We will look at two methods of randomly selecting numbers to choose a sample of 40 when all the members of a population of 200 have been previously numbered 1 – 200.

Random numbers generated by a calculator

To generate random numbers on the calculator we use the RAN button.

In this particular case we want to generate numbers up to 200 therefore we type:

200 Shift RAN

A number is generated and the corresponding numbered member of the population is selected for the sample. If a decimal is generated it will need to be rounded.

We continue with this method until we have a sample of 40.

Random numbers table

16405 82950 30197 64500 91130 58106 26415 72358
 80937 54607 31138 68716 83922 27129 28755 41225
 16405 82950 30197 64500 81200 98166 24435 62318
 10538 64917 51678 28736 43424 57328 88137 14201
 82551 81560 34246 15420 23452 78109 54010

To select our sample using a random numbers table like the one shown above, we read the digits in groups of three to generate 3-digit numbers (because the population is a 3-digit number).

Looking at the first line **16405 82950 30197 64500 91130 58106** ...

The numbers generated are 164, 058, 295, 030, 197, 645, 009, 113, 058, 106 ...

We ignore 295 and 645 as they are larger than 200 (the size of the population). 58 is repeated so we only include this the once.

The sample would therefore include the members numbered 164, 58, 30, 197, 9, 113 and 106.

We continue reading the digits in groups of three until we have our sample of 40.

Stratified random sampling E.g. The table shows the number of students of each nationality studying at The Welsh International College.

Nationality	American	British	Chinese	French	Spanish
Number of students	29	48	31	18	26

The Dean of the college would like to carry out a survey on student wellbeing. A stratified random sample of 30 students is to be used. How many students of each nationality will be included in the sample?

First we find the size of the population, i.e., the number of students at the college:
 $29 + 48 + 31 + 18 + 26 = 152$

We then use: $\text{No. selected from group} = \frac{\text{size of group}}{\text{size of population}} \times \text{size of sample}$
 to find the number of students of each nationality that needs to be in the sample.

$$\begin{aligned} \text{No. of American students} &= \frac{29}{152} \times 30 \\ &= 6 \text{ (to nearest person)} \end{aligned}$$

$$\begin{aligned} \text{No. of French students} &= \frac{18}{152} \times 30 \\ &= 4 \text{ (to nearest person)} \end{aligned}$$

$$\begin{aligned} \text{No. of British students} &= \frac{48}{152} \times 30 \\ &= 9 \text{ (to nearest person)} \end{aligned}$$

$$\begin{aligned} \text{No. of Spanish students} &= \frac{26}{152} \times 30 \\ &= 5 \text{ (to nearest person)} \end{aligned}$$

$$\begin{aligned} \text{No. of Chinese students} &= \frac{31}{152} \times 30 \\ &= 6 \text{ (to nearest person)} \end{aligned}$$

Check that the number for each group add to give the sample size needed, in this case 30. It's possible to have 1 more or 1 less then needed. This will be due to rounding the answer to the nearest person, therefore you may have to adjust an answer so that it fits with the sample size.