# M1.5 – Understand the principles of sampling as applied to scientific data

### Tutorials

Learners may be tested on their ability to:

* analyse random data collected by an appropriate means, e.g. use Simpson’s index of diversity to calculate the biodiversity of a habitat.

### Sampling

Sampling is an important part of scientific work: collecting data so that we can make inferences about relationships and processes in the natural world. For example, the only way to know the true mean height of a population of *Sequoia* (redwood) trees would be to measure every individual in the population. In reality this is practically impossible, because of the time it would take to do this, so instead we collect enough measurements in a **sample** to give us an accurate **estimate** of the **true mean** for the height of *Sequoia* trees, and we would refer to this as our sample mean.

There are broadly two ways to take samples: **randomly** or **non-randomly**.

**Random sampling** - When we carry out random sampling, the positions or sampling strategy used should be generated randomly. You as the experimenter do not choose which sites to sample, these are generated by a random number generator (such as the one on your calculator). In this way you do not influence the way in which measurements are taken.

**Non-random sampling** - A systematic, or even, approach to sampling. Measuring bacterial growth every 10 minutes, or taking a sample every 5 m across a woodland. This can be useful for measuring how things change over time or distances, where random sampling could cause strange clusters in sampling.

Stratified sampling – ‘Stratified’ or ‘layered’ sampling is a slightly more complex approach to sampling. The sampling approach is split according to predefined layers in the system being studied that could affect the results, for example splitting measurements of height by gender, and making sure that each ‘layer’ has a representative number of samples, in this case 50:50 to represent equal numbers of men and women.

In another example we might look at the distribution of invertebrates in a field, 10% of the field is water-logged and so I make sure that 10% of the samples are collected from water-logged areas.

When choosing which approach to use:

**Random sampling** is a good default when we have no reason to space things out evenly. An even distribution is not the same as a random distribution!

**Non-random sampling** is best applied when we have a good reason for wanting to know how things might change at regular intervals (time or distance are the most common uses
for this.)

When employing **stratified sampling** we should have a good reason for why it is reasonable to split the data collection according to the proposed ‘layers’. Within each ‘layer’ of our stratified sampling we may choose to sample randomly or non-randomly, whichever is most appropriate for the question being asked.

### Simpson’s Index

Sampling in field ecology is often used to measure abundance of different organisms in a habitat.

Once the abundance of all the organisms present in a habitat has been determined, scientists will often mathematically calculate the biodiversity present in a habitat. This can be done using Simpson’s Index of Diversity (D):

D = diversity index

N = total number of organisms in the ecosystem

n = number of individuals of each species

$$D= 1-∑(\frac{n}{N})^{2}$$

Simpson’s Index of Diversity always results in a value between 0 and 1, where 0 represents no diversity and a value of 1 represents infinite diversity. The higher the value of Simpson’s Index of Diversity, the more diverse the habitat.

### Produced in collaboration with the University of East Anglia

We’d like to know your view on the resources we produce. By clicking on ‘Like’ or ‘Dislike’ you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click ‘Send’. Thank you.

If you do not currently offer this OCR qualification but would like to do so, please complete the Expression of Interest Form which can be found here: [www.ocr.org.uk/expression-of-interest](http://www.ocr.org.uk/expression-of-interest)

Looking for a resource? There is now a quick and easy search tool to help find free resources for your qualification:
[www.ocr.org.uk/i-want-to/find-resources/](http://www.ocr.org.uk/i-want-to/find-resources/)

**OCR Resources**: *the small print*OCR’s resources are provided to support the delivery of OCR qualifications, but in no way constitute an endorsed teaching method that is required by the Board, and the decision to use them lies with the individual teacher. Whilst every effort is made to ensure the accuracy of the content, OCR cannot be held responsible for any errors or omissions within these resources.
© OCR 2017 - This resource may be freely copied and distributed, as long as the OCR logo and this message remain intact and OCR is acknowledged as the originator of this work.

OCR acknowledges the use of the following content: n/a

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: resources.feedback@ocr.org.uk

This resource has been produced as part of our free A Level teaching and learning support package. All the A Level teaching and learning resources, including delivery guides, topic exploration packs, lesson elements and more are available on the qualification webpages.

 If you are looking for examination practice materials, you can find the Sample Assessment Materials (SAMs) on the qualification webpages: [Biology A](http://www.ocr.org.uk/qualifications/as-a-level-gce-biology-a-h020-h420-from-2015/) / [Biology B](http://www.ocr.org.uk/qualifications/as-a-level-gce-biology-b-advancing-biology-h022-h422-from-2015/)