What is sampling?

- A **shortcut** method for investigating a **whole population**
- Data is gathered on a **small part** of the whole parent population or sampling frame, and used to inform what the whole picture is like

Why sample?

- In reality there is simply **not enough; time, energy, money, labour/man power, equipment, access** to suitable sites to **measure** every single item or site within the parent population or whole sampling frame.
- Therefore an **appropriate sampling strategy** is adopted to obtain a **representative**, and **statistically valid** sample of the whole.

Sampling considerations

- **Larger sample sizes** are more **accurate** representations of the whole
- The sample size chosen is a balance between obtaining a **statistically valid representation**, and the time, energy, money, labour, equipment and access available
- A sampling strategy made with the minimum of bias is the most statistically valid
- Most approaches ***assume*** that the parent population has a **normal distribution** where most items or individuals clustered close to the mean, with few extremes
- A 95% probability or **confidence level** is usually assumed, for example 95% of items or individuals will be within plus or minus two standard deviations from the mean
- This also means that up to five per cent may lie outside of this - **sampling, no matter how good can only ever be claimed to be a very close estimate**

Sampling techniques

Three main types of sampling strategy:

1. Random
2. Systematic
3. Stratified

Within these types, you may then decide on a; point, line, area method.

1. Random sampling

1. Random sampling

- Least biased of all sampling techniques, there is **no subjectivity** - each member of the total population has an **equal chance** of being selected
- Can be obtained using random number tables
- Microsoft Excel has a function to produce random number

Methodology

A. Random point sampling

A grid is drawn over a map of the study area

Random number tables are used to obtain coordinates/grid references for the points

Sampling takes place as feasibly close to these points as possible

B. Random line sampling

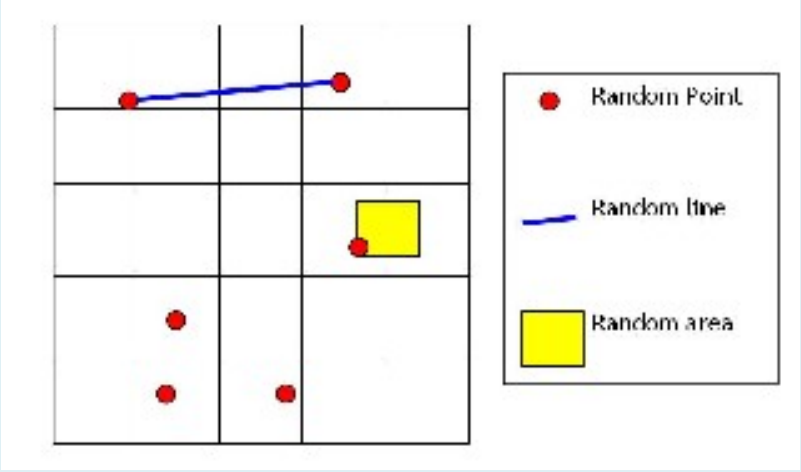
Pairs of coordinates or grid references are obtained using random number tables, and marked on a map of the study area

These are joined to form lines to be sampled

C. Random area sampling

Random number tables generate coordinates or grid references which are used to mark the bottom left (south west) corner of quadrats or grid squares to be sampled

Figure 1: A random number grid showing methods of generating random numbers, lines and areas



Advantages and disadvantages of random sampling

Advantages:

- Can be used with large sample populations
- Avoids bias

Disadvantages:

- Can lead to poor representation of the overall parent population or area if large areas are not hit by the random numbers generated. This is made worse if the study area is very large
- There may be practical constraints in terms of time available and access to certain parts of the study area

2. Systematic sampling

2. Systematic sampling

Samples are chosen in a systematic, or regular way.

- They are evenly/regularly distributed in a spatial context, for example every two metres along a transect line
- They can be at equal/regular intervals in a temporal context, for example every half hour or at set times of the day
- They can be regularly numbered, for example every 10th house or person

Methodology

A. Systematic point sampling

- A grid can be used and the points can be at the intersections of the grid lines (A), or in the middle of each grid square (B). Sampling is done at the nearest feasible place. Along a transect line, sampling points for vegetation/pebble data collection could be identified systematically, for example every two metres or every 10th pebble

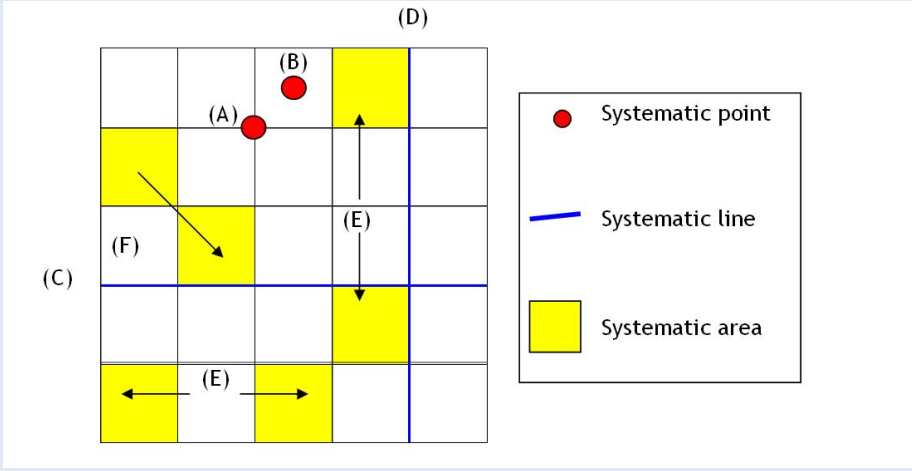
B. Systematic line sampling

- The eastings or northings of the grid on a map can be used to identify transect lines (C and D) Alternatively, along a beach it could be decided that a transect up the beach will be conducted every 20 metres along the length of the beach

C. Systematic area sampling

- A 'pattern' of grid squares to be sampled can be identified using a map of the study area, for example every second/third grid square down or across the area (E) - the south west corner will then mark the corner of a quadrat. Patterns can be any shape or direction as long as they are regular (F)

Figure 2: Systemic sampling grid showing methods of generating systemic points, lines and areas



Advantages and disadvantages of systematic sampling

Advantages:

- It is more straight-forward than random sampling
- A grid doesn't necessarily have to be used, sampling just has to be at uniform intervals
- A good coverage of the study area can be more easily achieved than using random sampling

Disadvantages:

- It is more biased, as not all members or points have an equal chance of being selected
- It may therefore lead to over or under representation of a particular pattern

3. Stratified sampling

Stratified sampling

This method is used when the parent population or sampling frame is made up of sub-sets of known size. These sub-sets make up different proportions of the total, and therefore sampling should be stratified to ensure that results are proportional and representative of the whole.

A. Stratified systematic sampling

The population can be divided into known groups, and each group sampled using a systematic approach. The number sampled in each group should be in proportion to its known size in the parent population.

For example: the make-up of different social groups in the population of a town can be obtained, and then the number of questionnaires carried out in different parts of the town can be stratified in line with this information. A systematic approach can still be used by asking every fifth person.

B. Stratified random sampling

A wide range of data and fieldwork situations can lend themselves to this approach - wherever there are two study areas being compared, for example two woodlands, river catchments, rock types or a population with sub-sets of known size, for example woodland with distinctly different habitats.

Random point, line or area techniques can be used as long as the number of measurements taken is in proportion to the size of the whole.

For example: if an area of woodland was the study site, there would likely be different types of habitat (sub-sets) within it. Random sampling may altogether 'miss' one or more of these.

Stratified sampling would take into account the proportional area of each habitat type within the woodland and then each could be sampled accordingly; if 20 samples were to be taken in the woodland as a whole, and it was found that a shrubby clearing accounted for 10% of the total area, two samples would need to be taken within the clearing. The sample points could still be identified randomly (A) or systematically (B) within each separate area of woodland.

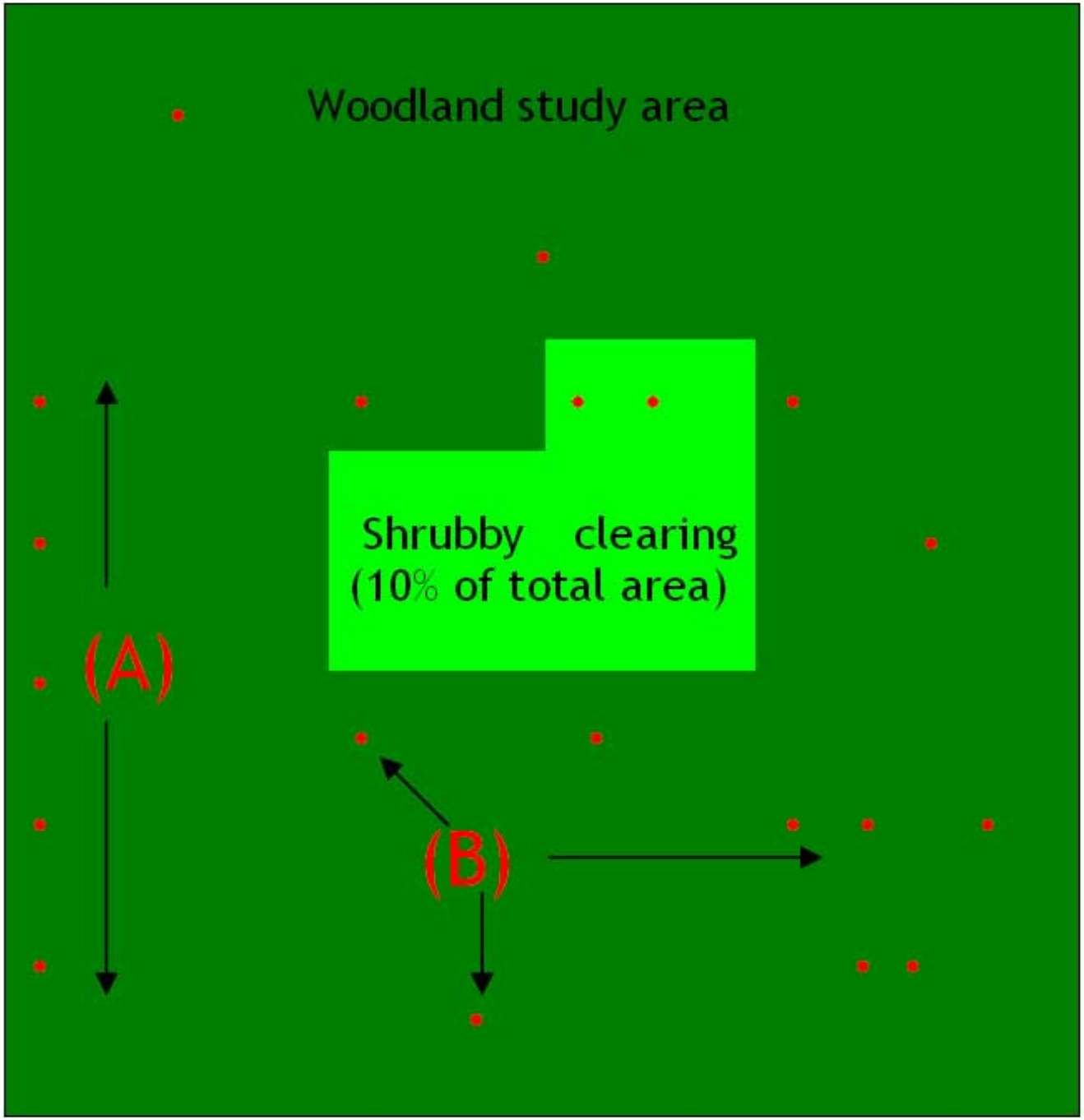


Figure three: A diagram highlighting the benefits of using stratified random sampling and stratified systemic sampling within certain fieldwork sites.

Advantages and disadvantages of stratified sampling

Advantages:

- It can be used with random or systematic sampling, and with point, line or area techniques
- If the proportions of the sub-sets are known, it can generate results which are more representative of the whole population
- It is very flexible and applicable to many geographical enquiries

Correlations and comparisons can be made between sub-sets

Disadvantages:

- The proportions of the sub-sets must be known and accurate if it is to work properly
- It can be hard to stratify questionnaire data collection, accurate up to date population data may not be available and it may be hard to identify people's age or social background effectively