1

Preparation

For many students, honing their research skills is an important component of their academic development. However, inexperienced researchers can be naïve in their approach, and may even attempt highly complicated studies that have little chance of being completed in the time and with the resources available. This chapter describes the thought and preparation needed to plan your project, particularly how to formulate your ideas into something structured and workable before going out into the field. Your research will search for explanations or relationships; make comparisons, predictions and generalisations; and formulate theories. Research is not simply an exercise in information gathering; rather, research is about asking questions that go beyond description and require analysis. Your research will be highly individual, and there are no set outcomes. You will form your own opinion, even if this disagrees with previous work. This is because progress in science results from the continual testing, review and criticism of other researchers' work. Do not expect your research project to answer all your original questions: it is much more typical to find that research generates more questions than it answers. Research submitted for publication or for examination should show evidence of originality. Even if your research is not wholly original, it can show evidence of original thinking. Although the prospect of carrying out original research may seem rather daunting, providing you do not exactly copy someone else's experimental design, methods, sites, etc., your research is almost certainly going to be original. There are several ways in which your work can be original:

- Executing an entirely new piece of work (e.g. studying a plant or animal for which there is little or no information currently available).

Practical Field Ecology: A Project Guide, First Edition. C. Philip Wheater, James R. Bell and Penny A. Cook.

- Adding knowledge in a way that has not been previously done before; many empirical studies do not develop new topics to study but instead angle their work with the use of original experimental designs, new statistical methods, etc. (for example, new insights might be generated from exploring the ecology of an otherwise well-studied animal at different sites to see whether food preferences differ between locations).

- Showing originality by testing somebody else's idea, or by carrying out an established idea in a new area, new experimental subject, etc., or by using existing data to develop new interpretations.

- Continuing an existing piece of work usually at your university or with a partner institution; for example, there are many long-term experiments that invite students to participate in summer work. These opportunities can be symbiotic and provide both you and the scientist running the project with more data that could elucidate a mechanism or generate new hypotheses.

- Originality may only be apparent in the breath of the study. Increasingly popular is 'cross-disciplinary' science, where, for example, soil scientists, botanists and entomologists converge on a subject matter or site and work together to test an overarching hypothesis.

All research needs careful planning (whether in the field or not). It is perhaps self evident that such planning should involve the correct use of equipment and choice of appropriate sampling methods and collection sites. In addition, a wide range of associated logistic, legal, and health and safety implications are also very important. Although many of these issues are equally important in field or laboratory-based investigations, field research may be more limited by time and other factors (access to sites, time of year, weather conditions) than research based entirely in the laboratory. Thus field research may need more careful consideration prior to implementation. Chapter 1 details some of the issues involved in planning and designing field research, and culminates in a checklist that may help to prevent problems once research is implemented. Chapter 2 deals with the techniques required for monitoring sampling sites and measuring physicochemical factors. Chapter 3 covers the methods used to sample static or relatively immobile organisms, and Chapter 4 extends this to studying mobile animals. Chapter 4 includes a consideration of monitoring behaviour and of dealing with both direct and indirect observations, as well as covering trapping and marking individual animals. In Chapter 5 we summarise a large number of different approaches to the statistical analysis of ecological data. Finally, in Chapter 6 we cover how to present your results and write appropriate reports.

Choosing a topic for study

The first stage of a research project is choosing a subject area on which to research (see Box 1.1 for a list of some texts that include ecological project ideas). As you will be devoting substantial time to your project, it is important to choose a topic that interests you. You may also wish to make your research relevant to your current or future employment. Pick a topic of the right size: neither too big nor too small. Looking at successful previous projects may assist you in judging how much can be done in the available time (ask those more experienced for examples of good projects to look at). Finally, your proposed project has to be feasible, for example in terms of equipment, access to sites and time-scale. Once you have selected your subject and provisional title, be prepared to be flexible and, if necessary, to change direction. This may happen for a variety of reasons, for example if a pilot study reveals a more interesting avenue for research or if your original ideas turn out to be unfeasible. You should note that the planning process should involve a consideration of the whole project to enable you to identify and deal with any potential problems before they become major issues (Figure 1.1). In all aspects, reading around the subject will allow you to use appropriate techniques, build on existing knowledge and avoid reinventing the wheel. Inevitably there can be logistical problems that influence your choice of site, or species, or otherwise prevent you from proceeding exactly as you would have wished. Although you can avoid such problems by careful planning, there are some aspects that you will not think about until you implement the research. A pilot study will help to identify such issues and may allow you to refine the study in advance of full implementation.

Box 1.1 Some sources of ecology projects

There are many resources that give examples of feasible ecological research projects. The series listed below are examples of some of those that cover either a wide range of habitat types or a range of organisms.

The Practical Ecology Series provide project ideas associated with grasslands (Brodie 1985), freshwaters (Gee 1986), the seashore (Jenkins 1983) and urban areas (Smith 1984).

Routledge Habitat Guides each include a section (Section 5) giving project ideas for the habitats associated with grasslands (Price 2003), uplands (Fielding and Howarth 1999), urban habitats (Wheater 1999) and woodlands (Read and Frater 1999).

The Naturalists' Handbooks (Richmond Publishing Company, Slough) contain many ideas related to studying a group of species (e.g. Gilbert 1993 on hoverflies or Majerus and Kearns 1989 on ladybirds), or different habitats (e.g. Hayward 1994 on sandy shores or Wheater and Read 1996 on animals under logs and stones), or implementing different techniques (e.g. Unwin and Corbet 1991 look at insects and microclimate, whilst Richardson 1992 examines pollution monitoring using lichens).

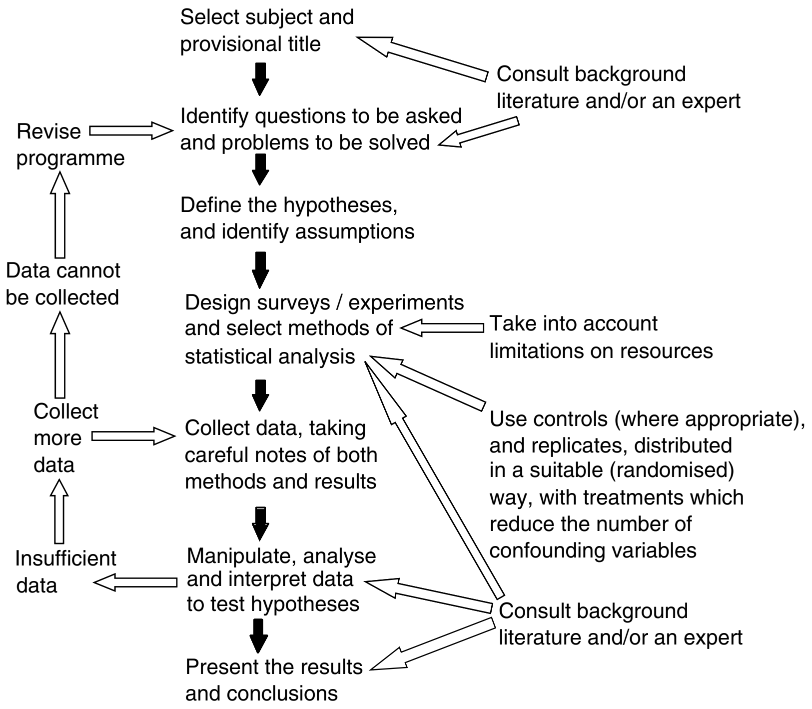


Figure 1.1 Flowchart of the planning considerations for research projects

Ecological research questions

Having decided on a provisional topic, the next step in the successful planning of any research project is to identify those questions you wish to ask and then to formulate the aims and objectives. There are various reasons for researching particular plants, animals or environments and this section provides a quick overview of the scope of ecological projects.

Many studies involve monitoring the number of species, number of individuals (relative abundance), estimates of population size or density (absolute abundance) or community structure (diversity, evenness and richness). Additionally, studies on animals may require observations of the behaviour of individuals or groups and their interactions with each other and their environment.

Monitoring individual species and groups of species

Sometimes, projects may be targeted at a single species. For example, where an important species is present because of its positive interactions (including any conservation or commercial value) you might require information about its distribution, population size and dynamics, age structure, behaviour, etc. Where you have a more negative view of a species (e.g. because it spreads disease, competes with native fauna and flora or is an invasive species that dominates a habitat to the exclusion of other species), you may need information about its distribution, dispersal ability, vulnerability to disturbance and predation, etc. An interesting aspect of biogeographical study is the examination of species distributions where species are expanding or contracting their ranges, perhaps as a result of climate change or other factors (either natural or human-influenced, e.g. habitat disturbance and fragmentation). Conversely, you may be interested in groups of organisms, examining the diversity of communities, the interrelationships between plants and animals in protected areas, or in establishing ecosystem function in relation to environmental legislation (e.g. the EU Water Framework Directive). Studies spanning a wide range of different taxa can be particularly valuable in understanding complex environmental systems, although they may be difficult to implement and the subsequent analysis and interpretation of the results can be complicated.

Monitoring species richness

In studies examining species richness, you might be interested in the presence or absence of one or more species (or other taxonomic group) in order to investigate the links between such species and aspects of the environment (perhaps in terms of the ecology of the species concerned, or in studies of pollution where the species may be useful as a biological indicator of certain toxins). Here, simply listing the plants and animals present may suffice. Although this may appear to be a quite simple approach, care needs to be taken to ensure that sampling techniques are used that are appropriate to both the organisms under consideration and the habitats in which they are found. For example, studies on the richness of bird species in urban parks may be complicated if some parks are dominated by relatively open habitats of amenity grassland and formal flower gardens, whereas others feature dense shrubberies and even woodland. Observations of the species present may be easier in the open habitats than under dense canopy. Care will thus be needed to ensure that all species are counted in an appropriate way at all sites. For these reasons, issues around surveying habitats and sampling organisms are considered in the next three chapters.

Monitoring population sizes and density

In population and density studies, it is the number of plants or animals of particular species that is of importance. Such studies may look at the number per unit area (i.e. the density) or calculate estimates of population sizes. Densities are taken from the estimated population divided by the area sampled. However, for mobile organisms it may be difficult to identify the spatial limits of the population (e.g. in studies of butterflies in agricultural sites some species may be highly mobile with individuals not being restricted to defined small sites). Under such circumstances, densities may be less useful than population estimations of the animals using particular sites. If populations of several species are being studied then it is important to ensure that the sampling methods used are appropriate to all the species being monitored. For example, in rainforests, some species of butterflies are found mainly within the canopy and are only occasionally caught at ground level and, conversely, some are predominant at ground level. Clearly, any survey comparing such study sites should incorporate sampling at both levels.

Monitoring community structure

Other studies might involve establishing the structure of the community of a specified area or habitat type (e.g. the community of fish in a lake, or the community of insects inhabiting a certain species of tree). Such studies may involve sampling a large range of quite different organisms. These may differ in size, distribution (both spatially and temporally), use of micro-habitats and, in the case of many animals, mobility. As such care needs to be taken to ensure that the methods are as comprehensive as possible and are not biased towards or against any particular species or groups of species. For example, sieving soil to examine the communities of animals living within different layers (leaf litter, humus layer, A horizon of the soil, etc.) may underestimate larger animals that are found at low densities (e.g. large ground beetles), and may overestimate species that are found in large aggregations if sampling happens to coincide with these groupings (e.g. some woodlice). Several different techniques may need to be used together during a single study in order to obtain a broad understanding of the community structure of such habitats.

Monitoring behaviour

Studies on animals may involve monitoring the behaviour of individuals, even if this is not the primary purpose of the study. Knowing whether rabbits are feeding, being vigilant for predators, etc., may be useful if numbers are being counted in particular sites. Of course, other research projects will focus primarily on animal behaviour. Such behavioural studies may involve the observation of a number of individual animals in a variety of settings, or the interactions that animals have with others of the same, and/or different, species. It is essential that the location and methods used by the observer do not influence the behaviours being monitored. For example, working too close to large mammals with young may be dangerous to the researcher and may mean that the major behaviour monitored is vigilance directed against the observer.

A note of caution

Although focusing in on the main aim of the research will help to formulate the procedure to be followed, you will also need to understand the limitations of the approach that you take. Census methods (e.g. simple species counts) can be quick to implement and provide substantial amounts of data in a short time. In contrast, techniques to assess population sizes or community structure tend to be much more time-consuming and may produce complex datasets. However, you should be aware that although it is usually possible to extract census information from population or community study datasets (albeit with a loss of detailed information), it is not possible to use census methods to assess community structure or population levels. In general, it is important to have at least some knowledge about the ecology and behaviour of the species or group of species under investigation when designing the research project, irrespective of the type of study being undertaken.

Creating aims, objectives and hypotheses

Once a topic for research has been chosen, you can work out the aims of the study. These are important since tightly defining the aims helps to focus more clearly on the work in hand and can avoid problems in implementation. Woolly aims such as ‘to investigate invertebrates under logs’ may be a starting point for a more focused aim such as ‘to determine whether the number of invertebrates found under logs is related to the size of the log’. This then leads to further questions, including:

- Which invertebrates are to be examined, that is should they be identified to species, or merely counted *en masse* or allocated to ecological groups (e.g. predators, herbivores, etc.).

- What is a log (i.e. when is a fallen piece of wood a log rather than a twig?) and how many logs should be investigated.

- How should we standardise or otherwise account for the condition and type of the logs (degree of decomposition, species of tree, etc.).

- Which measurements of size should be incorporated (e.g. length, width, surface area touching the ground, volume).

- Where should we sample the logs.

- Which statistical method(s) should we use to analyse the data?

Once these questions have been answered, they become objectives that can be used to determine the methods. The aims and objectives lead us to the setting up of working hypotheses. For example in our study of possible relationships between log size and the numbers of invertebrates found beneath them, we would set up a hypothesis to be tested. It is common practice that the hypothesis to be tested is a null hypothesis; in this context that ‘there is no relationship between log size and the number of invertebrate animals found underneath them’. Most univariate statistical tests examine the likelihood of the null hypothesis being true (see Chapter 5). A null hypothesis should meet the following criteria:

- Be a single, clear and testable statement – where more complex research questions are asked, you should break these hypotheses down into individual statements that are treated separately and tested in turn.
- Have an outcome, typically either ‘accept’ or ‘reject’ the null hypothesis.
- Be readily understandable to someone who is not a scientist.

Reviewing the literature

You should always review the planning and implementation of each stage of your research project by using current information, either from others who have been involved in similar research, or using texts, papers in journals or other information sources (e.g. the Internet), or a combination of these. Be aware of possible biases in the information used, especially where this is obtained from websites belonging to individuals (rather than respected organisations) that have not been independently validated. Most papers in reputable journals and many text books have been examined by independent referees, although even these may contain factual inaccuracies and personal opinions that may not conform to current opinion. Although considered the gold standard of information sources, even peer reviewed journals are subject to bias against the publication of negative results. It is important to start your review of the literature as early as possible, since it is an ongoing process throughout your research and should inform each stage of your project. At the very least you should begin by reading the literature to establish that your proposed idea has not been already published and to define the gaps in knowledge that you will attempt to fill. It is likely that as you read one paper, you will find references to other work that may be important.

If you are new to a subject matter, you should first try and locate seminal piece(s) of work in the field. Typically, this will be close to the top of a search list of highly cited papers and can be found by ordering a search by ‘times cited’. Take a detailed look at the seminal paper(s), the reference list and who is citing that paper. In journal databases (e.g. Web of Knowledge), citation networks can be viewed to examine the connectedness between a seminal paper and all those papers that cite it. This is useful because it can elucidate key papers in the field and reduce the search effort dramatically. Typically, your first search should include seminal works and a collection of the most recent papers in the field (i.e. last few years). It might be helpful to order these by journal impact factor (if available), since parochial journals may not contain as high-quality science, although sometimes smaller research papers with less apparent impact can provide valuable information in the form of species lists, new methods and negative findings that are often not reported in more mainstream journals. An additional word of warning: highly cited papers can also be poor papers in the field since other authors might simply be referencing them to make an example of that piece of work (e.g. ‘Black and White’s (2000) experimental design has been shown here and by others to be flawed’). Knowledge of the literature can assist in avoiding ‘blind alleys’ and unfruitful lines of enquiry or techniques. There are two main types of literature, primary and secondary.

Primary literature

This is first-hand information, for example, articles in specialist journals, reports, MSc and PhD theses. Journals that publish only refereed papers (i.e. those that have been through a peer review process) are the most important sources of primary, up-to-date information, and where possible your literature review should focus on this type of source. Other primary sources include technical reports, management plans, consultancy reports and species lists (e.g. from annual recorder reports). All of these can be useful sources of information for ecological projects, but you should be aware that they may not have been edited or their quality controlled.

Secondary literature

This is prepared from other sources of information, including textbooks, review articles, etc. If you are lucky there may be specialised books covering your subject area. These may provide a good starting point; however, since books are secondary sources of information, journals are a preferred source of reference for most research projects. The coverage in student textbooks is rather more superficial than that in specialist texts. If there are review articles on your subject, these may be useful to obtain an overview and as a source of new references. Reviews are found in edited book sections, journals that specialise in reviews (e.g. *Trends in Ecology and Evolution*[1]) and journals that have occasional review papers. Review papers in established scientific journals have usually been subject to peer review.

Other sources of information

Maps, personal notes, museum collections and archives may all contain information that can be useful in supporting your research. There are several software systems used within professional ecology that can be useful for research projects. These can help to set your work into a context grounded in practical conservation issues as well as supplying data on either a wider spatial or temporal basis. One such package, the Conservation Management System (CMS) software[2] is used by many countryside

---

[1] http://www.cell.com/trends/ecology-evolution/home

[2] http://www.cmsconsortium.org/software.html

managers to produce management plans, and Recorder[3] is the software used by the UK National Biodiversity Network to record, manipulate and map biological records.

Search terms

The use of targeted search terms can identify appropriate works and avoid too many articles or irrelevant papers from being selected. First, identify a list of key words or phrases that could be included in the title, abstract or key words of an article. Begin by using simple combinations of terms (e.g. woodland beetles, hedgerow birds). If you are finding too many papers, then either restrict your search to more recent publications or use more complicated combinations (e.g. by adding 'predation' to the key terms 'woodland beetles' you can significantly reduce the number of articles returned by a search). Note that most databases allow wildcard entries (e.g. an asterisk) to truncate key works (so that space, spaces and spatial can all be covered by spa*), although such terms can increase irrelevant returns (e.g. spa* will also cover any word beginning with spa: e.g. Spain, Spanish, etc.).

Reading papers

You will find more source papers than you have time to read. It is easy to get bogged down in the wealth of published material. Keep your subject area in mind and do not read everything indiscriminately. Skim-read a new reference to decide how much attention it deserves. Start by reading the abstract, skimming the subheadings and then the first paragraph or so of the introduction and the last paragraph of the discussion. Only read in detail those papers that are particularly relevant.

Keep a copy (photocopy, scanned image or electronic copy) of key references and make notes of (or highlight on photocopies) any useful information. Write down the full reference, since all material cited in a research report must be listed in full in the reference list; there are few things as annoying as having to refind the details of a reference that you read much earlier. For journals, the full reference includes the authors' names and initials, year, journal title, volume number and page numbers. For the other types of references see the guidelines for the reference list in Chapter 6. For books, in addition to noting the authors, publishers etc., take down the library classification number for your own reference in case you need to return to it at a later date. Increasingly, many papers are available on-line. The level of access depends on the services subscribed to by your organisation. All libraries offer an interlibrary loan service to provide access to works that are difficult to get. When using such a service, note the date on which you request any interlibrary loans to help to keep track of your requests.

---

[3] http://www.nbn.org.uk/getdoc/4bd40fa0-3692-4cea-9d81-701838bc8e87/Recording-software.aspx

Be critical as you read; do not accept everything as true just because it is published. Look at the evidence and decide whether the conclusions are justified, or whether the results could be interpreted differently. It is, unfortunately, not uncommon for assertions to be made with no supporting evidence. You will find that different authors in the same field may disagree. It is particularly important to distinguish opinions (and speculation) from evidence. You can make your own interpretations and conclusions from the work of others and cite them using expressions such as 'an alternative explanation for the results of Green and Brown (2010) is that…'. Read critically and keep your use of the information relevant for constructing your own account.

Practical considerations

Research (and especially field research) can be an unpredictable business. However, with careful thought it should be possible to ensure that most eventualities are covered. One of the major issues is the legal aspect (including access rights to land and the impacts on protected species or habitats). Health and safety is another obvious concern and it is essential that you ensure that there is no danger either to you or to those around you. In addition, practical approaches like effective time management, efficient data recording and security, and the appropriate use of equipment and techniques will also help to deliver a research project successfully.

Legal aspects

When planning fieldwork it is important to take into account your responsibilities and any legal implications of the work. At an early stage in the planning of the project, always seek permission to work on a site from the landowner and any other interested parties (e.g. relevant statutory bodies if the sites have some form of special protection, such as Sites of Special Scientific Interest or National Nature Reserves). Keep disturbance to a minimum, and remove as few plants or animals as possible. Identify specimens *in situ* (if you can) so that they need not be removed from the habitat. Whole plants should not be taken without the express permission of the landowner. It is good practice in the field not to pick plants of any kind, unless absolutely necessary. In many countries there are a number of protected species (e.g. orchids) that should not be uprooted, picked or harmed in any way. You should check with the appropriate governing organisation for the country involved for details on protected species. In some countries there is specific legislation covering protected species (e.g. the Endangered Species Act 1971 in the USA, the EU Birds Directive 1979, and the Wildlife and Countryside Act 1981 in the UK).

Some animals may not be disturbed or handled without a permit (e.g. birds and bats, among others, in the UK), nor may some microhabitats (e.g. badger setts in the UK).

Rare animals and plants are often protected by law and a licence may be required to handle or disturb them, without which you could face a large fine. You should also consider the ethical aspects of your study. This is particularly important where animals (especially vertebrates) may be harmed. Under these circumstances appropriate licensing authorities should be consulted. Even where this is not the case, the benefits of the study must be considered against the removal of individual animals and plants (sometimes in large quantities), possible damage to the environment, and other impacts including the spread of disease. Ethical aspects of work on animals are covered in work by Reed and Jennings (2007). Where relevant, your work may need to be examined through a local ethical review process. Most universities and similar organisations operate these, and further details can be found in RSPCA and LASA (2010). It is also worth reading the discussions by Minteer and Collins (2005a, b; 2008) and Parris et al. (2010) on this subject with reference to fieldwork. **Note:** this handbook is a guide that does not definitively outline the legal position or interpretation of any act or regulation. In all cases of protected species it is the responsibility of the researcher to check with the relevant bodies to understand what guidelines and regulations are in force.

Health and safety issues

Look after your own health and safety and that of those around you and try to avoid adversely influencing the environment. All investigations should be assessed for any risks, including those caused by the terrain, the techniques and any sudden changes in weather (see Barrow 2004). Any chemicals being used should be checked against appropriate regulations, and risk assessment should be produced to identify safe use, disposal and how to deal with spillage and accidents. In the UK such regulations (COSHH – Control of Substances Hazardous to Health) are covered by the Health and Safety Executive (HSE[4]). Many organisations have their own health and safety guidelines; in the absence of these, advice is available in Nichols (1999), Winser (2004) and Aldiss (2007). Pay particular attention when working at the coast (especially with regard to tides and hazards, including quicksand), rivers (above all with regard to potential flash floods), the uplands and mountain areas (especially with regard to sudden changes in weather conditions and the risk of exposure) and in situations where there is a risk from disease transmission, poisonous and venomous animals, or antisocial or violent behaviour from other people.

Try to avoid working alone in the field. If you must work alone, always carry a mobile (cell) phone and check out and back in with someone who knows your planned routine. Clothing and footwear should be suitable for the terrain and climatic conditions (warm and waterproof). Safety glasses and gloves should be worn to handle chemicals, and suitable gloves to protect you against thorns and infection from soil and

---

[4] http://www.hse.gov.uk/coshh/

water-borne disease. Keep your tetanus injections up to date, and take particular care where there is risk of Weil's disease (near to rivers and canals) and Lyme's disease (transmitted by ticks). Be aware of any other risks, including bites from venomous creatures (e.g. snakes), other toxic species (e.g. poisonous plants, scorpions, spiders and stonefish, which have poisonous spines), and the possibility of rabies from mammals. In general, ensure you are properly equipped (see Box 1.2) and avoid risks to help to ensure problem-free project work.

Box 1.2 Suggested minimum equipment required for field work

Always recommended

First aid kit
Map(s) of the area
Paper for recording (preferably in notebook form and waterproof if possible)
Pencils and sharpener (avoid ink if possible; even waterproof inks run when wet)
Mobile (cell) phone (fully charged and with spare batteries)
Whistle
Compass or global positioning system (GPS) receiver
Watch
Appropriate clothing and footwear
Appropriate safety equipment (e.g. gloves, safety glasses, etc.)
Appropriate sampling equipment (nets, traps, plastic tubes, plastic bags. NB: put any samples in a double plastic bag and label each bag so that if one label does come off, the other is there for reference)

Recommended depending on terrain, weather and timing and extent of work

Survival bag
Emergency food
Torch (fully charged and with spare batteries)

If working outside of your usual comfort zone, for example in locations overseas, check with researchers or organisations (e.g. The Royal Geographical Society of London) who have experience of working in such countries or habitat types. Even within your sphere of experience, avoid complacency since conditions may change and even small risks can be hazardous unless planning is comprehensive.

Implementation

A thorough literature search and any knowledge you have gained from other scientists experienced in the specific field of your proposed research is vital to define and refine the appropriate methods for your study. There will usually be a balance between the ideal solution in terms of the methods used, and logistical restrictions of time, and availability of equipment and expertise. Provided that your project has been well designed, and pilot studies have enabled you to refine your techniques, the implementation of the project should be straightforward. Here we emphasise the importance of careful note taking and time management during your project.

Equipment and technical support

Ensure the availability of equipment before starting, and obtain essential items well in advance of beginning your research project. You may need to allow adequate time to order specialist equipment or materials. If your project requires technical support, arrange this as far in advance as possible.

You need to be as familiar with your equipment as possible. This includes knowing how reliable it is likely to be under the conditions in which you are working and whether you need to have access to spare components or extra full items of equipment. For example, small mammals will eventually chew through the sides of aluminium small mammal traps (they rather more quickly get through the sides of equivalent plastic traps). Although it is possible to patch these up, this is tricky in the field and therefore spares should be taken. Anything that runs on batteries (e.g. data loggers or light traps) need to be recharged on a regular basis and spare batteries, bulbs, etc., should be available while in the field. If you are using multiple pieces of equipment, then you should ensure they are comparable. Different makes of bulb may provide different light levels in light traps, and monitoring equipment from different companies may have different levels of accuracy and resolution. Wherever possible, ensure that as similar as possible equipment is used for an individual project. Instrumentation errors may occur if users are unaware of the limits of the equipment (where attempts are made to estimate between gradations on an analogue scale). Make sure that you are familiar with all aspects of your equipment before engaging in field work. Calibration may also be important with equipment that requires regular calibration against standards of approximately similar values to the variables being measured (e.g. calibrating pH meters at pH 7 for neutral soil and water pH measurements). It is also important to take care of equipment, including protecting it against vandalism, theft and animal damage (many a moth trap has been trampled by inquisitive cattle when placed in their pasture).

Field/Laboratory notebook

Keep all your data and notes in an organised format, preferably in a hard-backed notebook (Box 1.3). Have a standardised way of recording your data, including everything that might be relevant: the date, weather conditions and notes of any important points that occur to you while carrying out the project. It is useful to record data in the same layout as you will on a computer spreadsheet for analysis (see p. 22). If you do use sheets of paper (similar to the one illustrated later in Figure 1.3), make sure that they all go into a ring binder as soon as possible. It is very easy for single data sheets to get lost. It is worth checking to see whether there is a standardised recording sheet available for use with the technique that you are employing. Biological Records Centre species recording sheets,[5] Breeding Bird Survey recording sheets[6] and freshwater invertebrate recording sheets[7] are some examples. Make photocopies of data at frequent intervals and scan them into a computer if possible. Where data loggers are being used either to note climatic variables (see Chapter 2) or to log behaviour (see Chapter 4), then make sure that you take backups of your files as soon as possible. Enter data and comments in electronic form whenever possible and create backup copies on a regular basis, including copies lodged on a networked drive or Internet hub.

Box 1.3 Keeping a field notebook

Use a field notebook to write down data, ideas, observations, tentative conclusions and hypotheses as you do your fieldwork to create an immediate and faithful history of your research. Produce comprehensive, clearly organised notes as a reference and so that you can reconstruct the research time-line and follow the development of your thoughts and ideas. Although you may use other collection sheets (e.g. pre-printed data collection forms to ensure data are collected consistently in different locations and at different times), your field notebook should provide the context for data collection and help resolve ambiguities or inconsistencies when preparing for analysis. After data analysis, reference to your notebook may generate further hypotheses and suggest further lines of enquiry.

Select an A5 or A6 hardback notebook with a spiral binding and wide-ruled lines, ideally on waterproof paper. Use a clutch-type propelling pencil with a moderately soft lead (HB or B). If you do not use waterproof paper, then encase your notebook in a plastic bag large enough to cover your hand and the notebook when writing. In very wet conditions write on an A4 sheet of white plastic with a thick soft pencil (use kitchen cleaner to erase your notes after transcription).

What should be recorded?

The first page should include contact details in case of loss, the subject of your research and the start and end dates of the period covered by that notebook. Include any conventions used, for example 'All times are recorded as local time'. Number the pages and ideally add a contents table to make searching for information easier. Write on the right hand page only so the left hand page can be used for ideas generated by reading about similar observations or relevant research papers. Leave a few lines between observations for comments to be inserted later (e.g. 'No bark damage here 23 June, see p39'). Add a 2 cm margin to write the time, location (e.g. from a GPS reading) or other identifying labels. Create lists of codes, acronyms, specialist terminology, etc., at the back and include any emergency numbers (e.g. those of field buddies). Other useful notes about equipment (how to use, limitations of instruments, etc.) and any numerical information you might require in the field (simple formulae for calculations, random numbers, etc.) can also be added here.

Before starting work each day, write down the date, weather, general location, nature of the habitat and purpose of the day's work. Write down any changes in weather or habitat that occur during the day, for example 'At 15.00 hours snow began to fall and visibility was reduced to 20 m'. When observing behaviour, note the sampling method, how animals were chosen for observation and the recording method (e.g. whether you noted all occurrences or used a time-sampled method). If animals or start times are chosen at random, note how this was done.

---

[5] http://www.brc.ac.uk/record_cards.asp

[6] http://www.bto.org/bbs/take_part/download-forms.htm

[7] http://www.fba.org.uk/recorders/publications_resources/sampling-protocols.html

Note the type and model number of any equipment (e.g. GPS receiver type Garmin 12). Some instruments need calibrating at intervals, so record the time of calibration and any raw data and subsequent calculations so that any arithmetic errors can be identified and corrected later. Use your notebook to create rough species accumulation curves, etc., so you can tell when you should stop collecting data (see Chapter 2). Along with observations, note the time and, if possible, the location from a GPS receiver. Although notes should be made at the time observations are made, it may be difficult to observe and write at the same time, but if you do rely on memory, you should note this. Write exactly what you see or hear, for example when describing behaviour do not ascribe a function to it in the guise of a description (i.e. do not write that a goose was 'vigilant' when you mean that the bird was in a standing posture with an elongated neck and raised head).

Sketches enhance any photographs you take of your study sites and you will have a sketch available in your notebook the next time you visit the area. Sketches can be added to subsequently (annotating any changes with the date of the amendment). The value of sketches can be increased by explanatory labels. A careful sketch can aid species identification and will help to jog your memory when you encounter a species in the future; such sketches are more valuable if labelled with the diagnostic feature(s) *you* use (e.g. 'two spots on forewing' or 'sepals reflexed'). Landscapes change over time and maps may not reflect this. In some cases no map of a suitable scale may be available and a sketch map can be made using compass and tape, or by pacing out distances using a pedometer. This may be adequate to note the locations of those animals or plants of interest.

It is also useful to record any notes and actions from supervisory or team meetings both as a reminder and to ensure that any designated actions have been completed as planned.

See also http://www.geos.ed.ac.uk/undergraduate/field/fnb

Time management

Conducting a piece of research within a relatively short time span is a demanding process, and will require careful time management. You will need to allow time to check the feasibility of the research project and should also add sufficient time for method training or familiarisation. Ensure that you leave time spare to allow for the almost inevitable problems associated with both field work and laboratory analysis. Check out your methods by first running a pilot to help identify any pitfalls and inadequacies. It is easy to underestimate how long it takes to analyse, interpret and write up a research report. Figure 1.2 illustrates a timetable (in the form of a Gantt chart) for a research project expected to last 33 weeks – around a full academic year: many research projects are substantially shorter than this. Note how the time allocated to actually collecting the data is relatively short and overlaps with the ongoing literature review and writing process. Also, you are strongly urged to begin writing the report before completing your data collection; it should be possible at this stage for you to write the methods and introduction since the former will describe what you have been doing, whereas the latter reflects your background reading of the literature. Table 1.1 shows an example time-scale for a research project lasting a week (e.g. on a field course). Note that even with such a short time-scale, sufficient time needs to be devoted to planning the project to ensure the best chance of a successful outcome (Box 1.4).

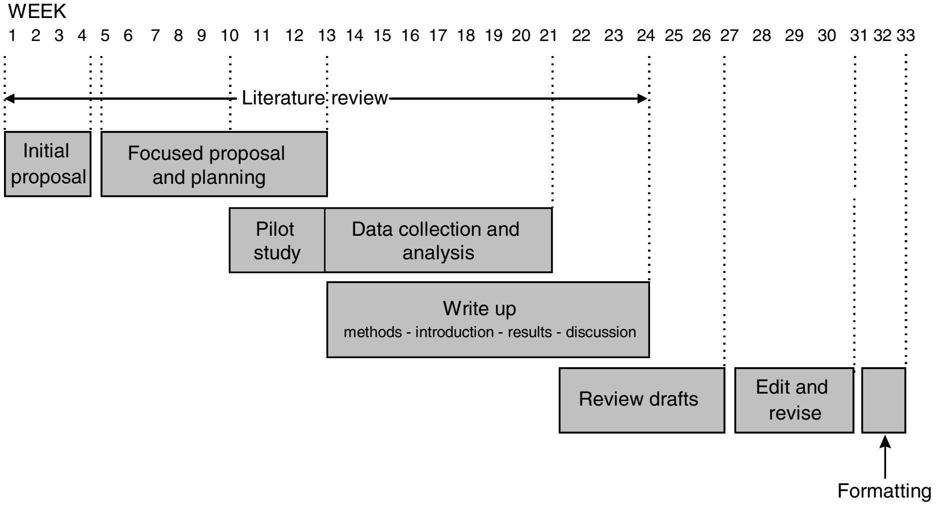


Figure 1.2 Example time-scales for a medium-term research project

Table 1.1 Example time-scales for a short research project

Day	Morning	Afternoon	Evening
1	Select topic, identify aims and hypotheses, create programme of study, select sites and techniques, identify major resources required, complete risk and ethical assessments	Test techniques, train on equipment, implement a pilot study	Evaluate pilot study, amend programme of study
2	Implement amended programme and collect data		Enter data into spreadsheet, write methods
3	Collect data		Enter data into spreadsheet, write introduction
4	Collect data		Enter data into spreadsheet, edit introduction and methods
5	Collect data		Enter data into spreadsheet, plan results tables and figures
6	Collect data		Analyse data
7	Write results section	Write discussion	Complete report

Box 1.4 Some tips on time management

- Be realistic and work within your strengths and weaknesses.
- Plan your long-term goals.
- Have a weekly plan, with realistic and achievable targets, and update this on a regular basis to reflect your progress.
- Identify not only the key phases of your research project but also other areas that will take up your time (both in terms of study and general living) to ensure that your research time-scales are realistic and achievable.
- Prioritise your work into that you have got to do, that you ought to do and that you would like to do but may not have time.
- Use odd snatches of time: a trip by train may be an ideal opportunity to read references or edit your manuscript.

Project design and data management

Since research is about asking questions, you need to design your project so that it will answer the question effectively, without allowing your design to introduce ambiguous results or results that are open to other interpretations. This is where the planning phase starts to define what you are going to measure and how. If we look at an example where we investigate the types of birds found inhabiting a woodland patch, then we have a choice of ways in which we record the data. We might note how many individual birds there are, or the numbers of each feeding type (insect feeders, seed feeders, etc.) or how many individuals there are in each species. These measurements enable us to obtain a picture of the birds found in a woodland patch. If we only monitor birds in a single woodland patch, we could worry that our chosen woodland is unusual in some way and therefore not representative of woodland patches in general. We could therefore examine a series of patches and obtain data for 10 or more. Now if we wish to describe how many birds were found in all of these woodlands, we require some sort of descriptive statistic to summarise the information across 10 or more patches. Descriptive techniques include estimates of the average values per sampling unit (e.g. per site), population estimates and densities, methods of describing distributions (i.e. whether organisms are distributed randomly, evenly or in aggregations) and measures of community richness, including diversity and evenness indices. These techniques are discussed in more detail in Chapter 5.

Most projects go beyond a simple description of particular species and sites in an attempt to make comparisons or generalisations that can hopefully have wider applicability. For example, if we decide to investigate whether the number of animals found under decaying logs on a woodland floor is influenced by the size of the log, we might approach this in one of three ways:

1. by looking at possible differences between samples, for example if the logs were easily divided into two classes (large and small), we could compare the numbers of animals found under each size class.
2. by looking at possible relationships between variables, for example we might have a wide range of sizes of logs and decide to examine whether the number of animals varies in some systematic way (either increasing or decreasing) as log size increases.
3. by looking at possible associations between frequency distributions, for example, we could compare the frequency of obtaining predators, herbivores, decomposers, etc., under each of two size classes of logs.

We will examine each type of question in a little more detail later in this chapter (p. 29), while the analytical techniques needed to answer these questions are described in Chapter 5. Other questions that might be asked include looking at the similarity of sites based on their species composition (i.e. are the animals found under logs from different tree species similar in species composition to each other) or predicting the presence or numbers of a species from a knowledge of the environmental conditions (i.e. is the presence of wood ants nests predictable if we know the woodland type, topography, microclimate, etc.).

We need to ensure that our study does not produce ambiguous results. For example, in a comparison of the invertebrate diversity between urban ponds and rural ponds we could aim to include the size of each pond studied into the survey design. If we did not manage this, and found that the rural ponds surveyed happened to be both larger and contain more invertebrates, it would not be clear whether the results were due to rural ponds being more diverse or whether it was simply an effect of pond size. The correct experimental design would be to either standardise on a given pond size for both environments, or to make sure that the full range of pond sizes were included in both environments (and measured, recorded and built into the subsequent analysis). Other factors that would have to be standardized, or at least recognised as covariates, in this particular study would be the quality of the water, the pH, age of pond, and so on.

The goal of the study may be to get a deeper understanding of the system by gathering a wide range of variables. In the pond survey example, this might mean that in addition to pond size, various measures of water quality and chemistry (nutrient status, oxygen content, pH, etc.) and the numbers of each species of plant and animal we observe. For a large number of ponds (each usually recorded as a single row in a data spreadsheet) there may be a large number of variables (usually recorded as columns) leading to a large data matrix. In order to examine and make sense of such a complex dataset, we would need to move into the realm of multivariate analysis (see Chapter 5).

Designing and setting up experiments and surveys

There are two main approaches to collecting data: experiments and surveys. An experiment involves the manipulation of a system, whereas a survey depends on observations being taken without manipulation. For example, if we were interested in how many invertebrates could be found under logs of varying size, we could either survey a woodland floor finding as many logs as possible and recording both the number of invertebrates and the size of the log, or we could devise an experiment where we placed logs of differing sizes on a woodland floor and after a period of time examined the number of invertebrates underneath them. The advantage of the experimental approach would be that we could standardise all aspects of the logs except for size, for example age, the degree of decay, the type of wood and the distance between logs. All of these factors may influence the invertebrates found and confuse any relationship with log size. However, with a survey we would get an impression of what was happening in a real life situation (i.e. under logs that had been naturally deposited). Moreover, we may decide that the experimental approach is damaging to the environment, here adding unnaturally deposited logs to a natural system. In addition, for practical reasons we might decide that the colonisation of newly introduced logs by invertebrates would take longer than the time available for the project to be completed. In most environmental research programmes, surveys are useful for generating ideas about important factors, but because of the additional complexity in real situations, cannot identify cause and effect. Because experiments strip away the additional complexity, they are more useful in identifying cause and effect, but less likely to be applicable to real life situations. When designing experiments it is important that as many factors as possible are kept constant. So, for example, if we are interested in identifying whether an increase in pesticide concentration will lead to a decrease in aphid infestation of a crop, then the same amount of water (assuming this is the solvent or carrier for the pesticide) should be used for each application (irrespective of the concentration applied) so that we are testing the amount of pesticide added, rather than the amount of water added. In addition, it is important where possible to include a control treatment. In this example we would use a water-only treatment to see if the addition of any water had an impact. If we did not do this and found a reduction in aphid numbers with any application of pesticide, we would be unable to tell whether this was due to the pesticide or the fluid added.

Types of data

In order to design an appropriate experiment or a survey, you need to think about the type of data you wish to collect. The pieces of information that are recorded (e.g. height of tree, number of birds, density of plants per unit area) are termed variables and may be in the form of one of three types of data. The simplest type is categorical or nominal data where each value is identified as one of several distinct categories (e.g. male or female animals; purple, red or yellow flowers; grasses, ferns, herbaceous plants, shrubs, trees). Where we can place the categories in some kind of logical order, so that the data are able to be ranked, this is called ordinal data (e.g. large, medium-sized or small ponds; above the high-tide line, mid-shore and below the low-tide line on a rocky shore). The most detailed types of data are those measurements that not only can be placed in a logical order, but where there is a known interval between adjacent items in the sequence (e.g. the number of deer in a herd; the temperature in the centre of patches of plants of differing sizes; the depths of a series of ponds). There are two types of measurement data: interval data and ratio data (Box 1.5). In most cases the analysis of interval and ratio data uses the same techniques and so in this text we will tend to combine them and refer to them as interval/ratio data or measurement data.

Box 1.5 Differences between interval and ratio data

Interval data have no true zero so that negative values are possible (as in temperature measured on the Celsius scale where 0 °C refers to the freezing point of water rather than the lowest possible temperature) and where measurements cannot be multiplied or divided to give meaningful answers (as in dates).

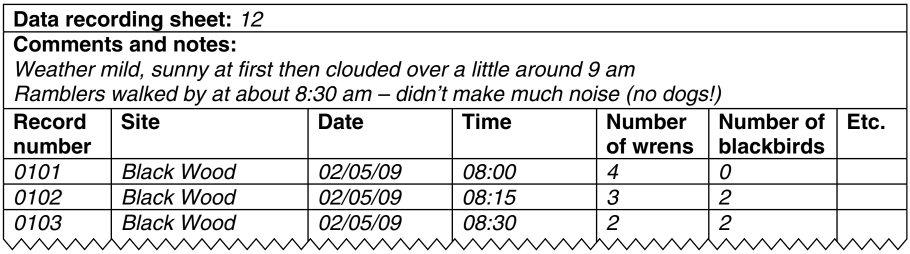
Ratio data are measurements that have an absolute zero point that is the lowest possible value (as in temperature measured on the Kelvin scale where 0 Kelvin is absolute zero) and so negative values are not possible (e.g. you cannot have –6 foxes). With ratio data all basic mathematical operations can be performed to give meaningful answers; e.g. you can derive a ratio of water lost from soil following drying out as follows (where the original mass $= 20\,\text{g}$, and dried mass $= 16.5\,\text{g}$):

the proportion lost on drying $= (20 - 16.5)/20 = 0.175$, i.e. 17.5%

Note that we can readily reduce measurement data to ordinal or categorical, but not the other way around. Thus, if we count the numbers of invertebrates of different species on a particular type of plant, we could subsequently express this in order of dominance from abundant through to rare (an ordinal scale) or indicate the presence or absence of different species (categories). However, if we originally merely record presence and absence of species, we cannot subsequently calculate the numbers of individuals. Thus, if in doubt, it is safest to collect the information at the highest resolution possible.

It is good practice to use a standardised data recording sheet (ideally in your field notebook) that is as similar as possible to the way in which data will be entered into a computer for analysis to avoid data transcription errors in moving from paper to a computer spreadsheet. In our example (Figure 1.3) we have two types of variables: fixed and measured. It is easier to deal with these in order so that fixed variables come first, followed by measured variables. Fixed variables are those determined by the research design and do not vary during the investigation (record number, site, day and time). Hence these can be added to the recording sheet early in its production. Measured variables on the other hand are those factors recorded during the investigation, values of which will vary depending on the site, day, time, etc. (numbers of wrens, blackbirds, etc.).

Sometimes derived variables are also required (i.e. variables produced from measured data, e.g. the proportions that each species forms of the whole catch). Such derived variables can be added to the right of the measured data once the latter have been entered on a computer spreadsheet, since the required computations are easily carried out using spreadsheet functions. In most cases, data will be recorded as numerical values. Where categories (e.g. site) occur, codes or names can be used, although some computer programs will not accept letter codes, so you may need to allocate numeric codes to such variables. You should make sure that any paper copies of results sheets are photocopied or scanned as soon as possible after completion, and that electronic copies are properly backed up.



Data recording sheet: *12*

Comments and notes:

Weather mild, sunny at first then clouded over a little around 9 am

Ramblers walked by at about 8:30 am – didn't make much noise (no dogs!)

Record number	Site	Date	Time	Number of wrens	Number of blackbirds	Etc.
0101	*Black Wood*	*02/05/09*	*08:00*	*4*	*0*	
0102	*Black Wood*	*02/05/09*	*08:15*	*3*	*2*	
0103	*Black Wood*	*02/05/09*	*08:30*	*2*	*2*	

Figure 1.3 Example of a data-recording sheet for an investigation into the distribution of woodland birds

Sampling designs

When implementing a project, it is rarely possible to collect information on all the animals or plants present. Usually we need to use a sample that we hope to be representative of the situation as a whole. The total number of data points that could theoretically be gathered is known as the population (this is a statistical population rather than the population of animals or plants – Box 1.6); the actual number of data points is termed the sample size. Larger samples are usually more representative of populations, although this depends on the variability of the system being studied (small samples may be reliable representations of populations with low variability). Those elements of a system that are calculated (e.g. the mean number of plants per square metre in a meadow) are termed statistics and are estimates of the true attributes of a statistical population (called parameters – Box 1.6). So if we counted all the plants in the entire meadow, we would be able to calculate the actual mean value per square metre (a parameter). Since it is usually impractical to count all individual plants, in reality we usually count plants in a subset of the meadow (i.e. take a sample) and calculate the mean numbers per square metre using this sample on the expectation that it will be representative of the whole site (a statistic). This sort of situation occurs in many types of survey. For example, market researchers obtain opinions from large groups (samples) of people and use these to indicate the attitudes of the population as a whole.

Box 1.6 Terms used in sampling theory

A **population** is a collection of individuals, normally defined by a given area at a given time. For example, scientists refer to the decline in the world population of Atlantic cod in the last century or the annual harvest of Northeast Atlantic cod. These are both true populations.

A **sample** is a term that can be used ambiguously but is a subset drawn from a population, which usually includes a quantity (e.g. 100 individual fish taken from the Northeast Atlantic cod population).

A **parameter** is a population metric that is estimated from a variable (e.g. the mean body size of Northeast Atlantic cod) and can be used to summarise data. Importantly, statistical tests aim to derive parameters from a population in order to test for differences, relationships, associations, etc.

A **variable** is a measurement that may change from sampling unit to sampling unit (e.g. the body size of Northeast Atlantic cod taken from a sample) and can be used to summarise collected data (e.g. by taking the mean).

Choosing the sample to be taken needs some care and at this point it is worth discussing why replication is important. Since environmental systems are usually intrinsically variable (i.e. physical, chemical and biological factors differ spatially and temporally), the larger the sample, then the more representative it will be of the population (i.e. the more of the natural variation will be covered). However, the larger the sample, the more time and effort it will take to collect it. There are methods to calculate the optimum sample size; however, these rely on knowledge of the variability of the system. This is rarely known in advance, although a small pilot study may give some indication. If it is known or suspected that there is substantial variability, then a large sample should be taken. In most ecological surveys a large sample would include over 50 data points. However, where the population is likely to be very large and variation is expected to be great, even larger sample sizes may be required. Otherwise it is best to aim for as large a sample as possible after taking into account constraints including the size of the workforce, the time available and how much material is present in the system under investigation. Where several levels of a number of variables are to be analysed (e.g. male and female animals of each of three different age groups: young, mature and old), then it is important to take sufficient replicates of each subgroup (e.g. young males, mature males, etc.) to be able to account for within-group variability. This will inevitably have an impact on the required sample size and is another reason why the intended statistical analyses should be considered at an early stage of project planning. See Krebs (1999), van Belle (2002), and various on-line calculators[8] for further details of the different calculations that can be used to estimate sample sizes depending on the intended statistical analysis technique to be used.

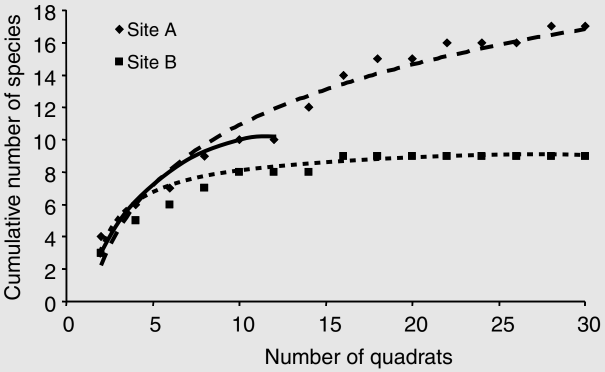
---

[8] For example, Rollin Brant at the University of British Columbia (http://www.stat.ubc.ca/~rollin/stats/ssize/index.html) and The Australian National Statistical Service (http://www.nss.gov.au/nss/home.nsf/pages/Sample + Size + Calculator + Description?OpenDocument)

In surveys of community structure, it may be important to know that the majority of species in an area have been recorded. In this case, species accumulation curves may help. At its simplest, this involves plotting the accumulated number of species against increasing sampling effort. Sampling effort is the number of sampling units (quadrats, pitfall traps, animals handled, hours of observations, sites surveyed, etc.). Box 1.7 illustrates the use of species accumulation curves in quadrat sampling (see Chapter 3). There are a variety of methods of modelling species accumulation curves (see Colwell, Mao and Chang 2004; Magurran 2004 for further information) and many standard software packages include routines for this (e.g. those obtained from Pisces Conservation[9]).

Box 1.7 Species accumulation curves for two sites

By plotting the cumulative number of species found against the number of quadrats examined, it can be seen that as the number of quadrats used increases, the number of species also increases. At the point at which the curve levels off towards horizontal (the asymptote), we may assume that we have obtained the maximum number of species and can stop sampling. For site A (dashed line, diamonds), we may not yet have reached the total number of species, even after 30 quadrats and should consider increasing the sampling effort. For site B (dotted line, squares), it appears that we have reached about the maximum number of species that we can expect to get. In fact, we probably reached this number at round about 16 or so quadrats. This difference between sites A and B might reflect not only a difference in the number of species found there, but also a difference in heterogeneity of the site, with site A being less homogeneous than site B. Note that had we looked at the data for site A after 12 quadrats (solid line), we might have assumed that we had reached the maximum number of species as the curve levels off. This highlights the importance of collecting past the initial point of curve levelling to check that it truly does reflect the asymptote.



---

[9] http://www.pisces-conservation.com/

Since you are generally taking a sample in order to make a valid estimate of a parameter of the population (e.g. the number of species, the mean temperature, the proportion of predators), a central requirement is that the individuals sampled are independent of each other. It is important to recognise (and avoid or account for) situations where the individuals sampled are linked in some way as a result of the sampling design. For example, we might compare the number of spangle galls found on leaves chosen at random on oak trees growing in clumps with those on isolated oak trees. If we found over 20 trees in separate clumps but only 10 isolated trees, we might be tempted to take double the measurements from the isolated trees. However, this would mean that individual data points from isolated trees were linked by virtue of the tree on which they were growing, and shared many different attributes with each other. Such data would not be independent of each other and hence may cause problems in interpretation since we would be unsure whether any differences between clumped and isolated trees were due to the multiple measurements from some trees. It would be better to use unbalanced sample sizes (i.e. 20 clumped and 10 isolated trees) than use non-independent data. Similarly we should not take data from more than one tree in any clump in case these are linked in some way. From a statistical analysis point of view, few tests require equal sample sizes, and even where this is a problem it would be preferable to reduce the number of trees from clumps that were measured.

If we survey a pond in order to look at the animals and their relationships with several physical, chemical and/or biological factors, then no matter how many replicates we take, we are merely describing what happens in a single entity (i.e. this one pond). Such a study does not tell us anything about pond ecology in general, and the use of such replicates is termed pseudoreplication and should be avoided (Hurlbert 1984, van Belle 2002). In order to broaden our approach and gain more of an understanding of ponds in general, we would need to study a large number of separate ponds. Thus, studies of single sites or small parts of sites may not reveal information applicable to the wider ecological context.

In some situations the data collected are linked to each other by design. For example, we might be interested in comparisons of matched data (e.g. examining the animals found on cabbages before and after the application of fertiliser or pesticide, or the numbers of mayfly larvae found above and below storm drain outflows into a series of streams). These designs can be perfectly sound, but because the data are matched (by cabbage or by stream) we require a slightly different approach to the resulting analysis (see Chapter 5).

When designing your sampling strategy, it is important to consider the variability and whether the timing or order of sampling might bias the result by measuring only part of the potential variation. For example, sampling the insects present on thistle flower heads will be biased if all the data are collected in the early morning since this will miss any animals that are active later in the day. If two areas are being compared, sampling one site early and one site later will introduce another variable into the comparison: we would not just be looking at the two sites, but also at two times of day. Since it would be impossible to separate the two variables, it would be difficult to draw conclusions from such a survey design. It is in managing some of this variability that experiments come into their own, because they standardise as far as possible the conditions under which the subjects are examined. It is much easier to design an experiment where only one factor (also known as the treatment) is manipulated, while all others remain constant. However, if we wished to survey a real life situation (as opposed to examining a rather more artificial experimental design) then we would take into account the time of day. We could do this by designing our survey so that we alternated the measurements or observations that we took from our two sites, sampling first one then the other, then back to the first, and so on, to get a spread of measurements for each site over the day. Alternatively, we could sample on successive days, reversing the order in which we sampled the sites on each day.

There are several sampling layouts that help us to avoid bias. One commonly used approach is random sampling. Here a random sequence is used to determine the order in which to sample plants, or the coordinates to sample experimental plots or survey sites. Hence, if we wanted to randomly sample 1 m × 1 m quadrats in a field, random coordinates can be used to position the sampling sites (Figure 1.4a) using pairs of random numbers generated using a calculator or computer, or obtained from a table (Table 1.2). This works by using pairs of numbers as sampling coordinates, so if we have coordinates of 23 and 85 in a sampling grid that is 10 m by 10 m, we would place our quadrats 2.3 m along the base and 8.5 m up the vertical axis. Random sampling may also be used to determine which site is visited first if sites are allocated number codes that are then selected randomly from the table.

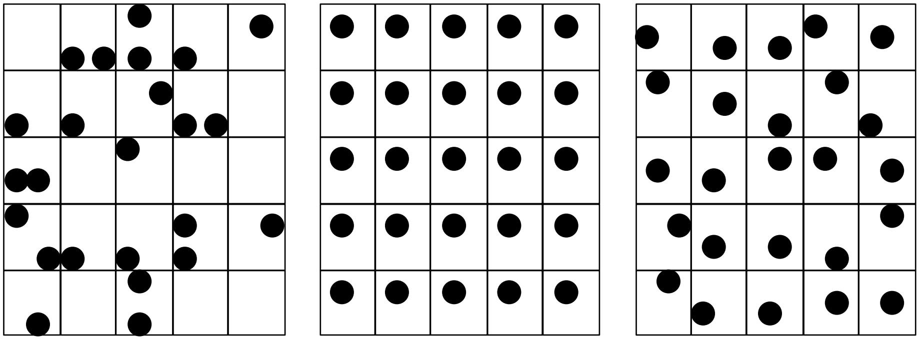

(a) random sampling (b) systematic sampling (c) stratified random sampling

Figure 1.4 Examples of sampling designs

Table 1.2 Random numbers

Coordinates can be extracted simply by taking pairs of random numbers in sequence from the table (e.g. 23, 85 – shaded values – provides the position within a sampling area where we would take the first measurement of a series).

23	85	56	84	92	4
62	51	27	74	83	84
56	32	87	75	95	5
87	7	20	30	25	12
99	86	29	41	29	39
31	73	30	73	27	97
24	38	91	16	17	66
94	59	12	17	37	39
41	67	25	42	2	84
32	67	48	99	74	3
68	1	59	20	25	7

Although random sampling is often appropriate for selecting sampling points, where there is a great deal of variation across a sampling unit, such as a site, by chance the coverage may not include all of the heterogeneity present. For example, in Figure 1.4a, the two squares in the lower right of the sampling site have no sampling points. If the site was reasonably homogeneous, then this would not be a problem. However, if these small squares represented the only damp area within the site (covering around 8% of the total area), then this particular variation would have been missed altogether. An alternative strategy would be to use systematic sampling (Figure 1.4b). This is an objective method of spreading the sampling points across the entire area, thus dealing with any spatial heterogeneity. So, to systematically sample the insects on trees, we might collect from every tenth tree in a plantation.

Usually systematic sampling would provide us with random individuals unless for some reason every tenth individual is more likely to share certain characteristics. Suppose we used systematic sampling to examine the distribution of ants' nests in a grassland. We could place 2 m by 2 m quadrats evenly 10 m apart across the site and then count the number of nests within each quadrat. However, if ants' nests are in competition with each other, they are likely to be spaced out. If this spacing happens to be at about 10 m distances, we would either overestimate the number of nests if our sequence of samples included the nests, or underestimate if we just missed including nests in each quadrat. It would be better in this situation to use a mixture of random and systematic sampling (called stratified random sampling – Figure 1.4c) where the area was divided into blocks (say of 10 m × 10 m) and then the 2 m by 2 m quadrats were placed randomly within each of these. This type of sampling design can also be applied to temporal situations by, for example dividing the day into blocks of 4 hours and allocating the order of the sites to be sampled within each block using different random numbers.

More sophisticated methods of laying out sampling plots (or allocating sampling periods) may be useful for experiments. One example is the use of a Latin square design (Figure 1.5) that ensures that experimental treatments are equally distributed across the rows and columns of a sampling design.

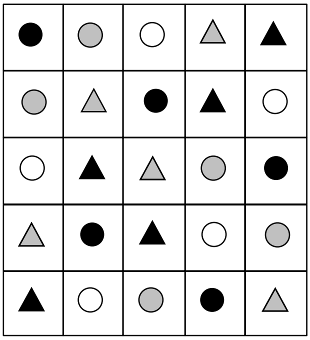


Figure 1.5 Latin square design for five different treatments
Each treatment is represented by a different symbol.

Planning statistical analysis

Although at this stage we will not discuss in detail the ways in which data are analysed, it is important to at least have sight of the likely methods that may be used. This is because different statistical methods are required to deal with different research questions. Most techniques require specific types of data to be gathered, and lack of care at this stage could result in data being collected that cannot answer the question posed. In addition, some statistical analyses have minimum numbers of data points that are needed to obtain meaningful results and a few need balanced designs (i.e. the same number of data points for each factor measured). In this section we will discuss some of the major types of analyses that you could employ to answer certain commonly asked types of questions. Further details are given in Chapter 5. As always it is worth looking at the literature to see what types of analysis have been used in similar studies to the one you propose to do. There are several major groups of analysis based on the broad types of approach required.

Describing data

We need a variety of techniques to describe the data that we collect. This might be as a data exploratory technique (to check the data to see how variable a dataset is, or what sort of distribution we get, etc.), to understand some aspects of the data (e.g. how diverse communities are) and for communication purposes (to be able to discuss the results, orally and in writing, with other people).

Here simple plotting of measured variables on frequency histograms (or tables), cross-tabulation of one (nominal) variable against another, and examining the range of the data (from the minimum to the maximum) may help to check for errors and ensure that we can chose the correct type of test for subsequent analysis. Extracting statistics, such as diversity indices (to describe species richness) or evenness (to describe how equal the proportion of species is within a community) can be important in being able to assess what sort of community we have. Likewise the estimation of population size or density can be important in some studies. Similarly, we can use the average value of a variable to describe the magnitude of the majority of the data points (usually in conjunction with some measure of how variable the values are and how many data points were collected). Chapter 5 reviews these descriptive statistical techniques and more detail can be found in Wheater and Cook (2000; 2003).

Asking questions about data

If we wish to ask specific questions of the data, then we are in the realm of inferential statistics. These usually involve the testing of hypotheses. It is standard practice to set up a null hypothesis alongside the questions to be asked. The null hypothesis tests the chance of there being no significant difference between samples (or relationship between variables, or association between categories of variables). So if we wish to know whether there is a difference between two samples (e.g. comparing the number of birds found in deciduous woodlands with the number found in coniferous woodlands), then we actually test the null hypothesis that *there is no significant difference between the number of birds in deciduous and coniferous woodlands.* Note that we are looking at ‘significant’ differences. These are differences that are unlikely to have resulted from random variation in the individual woodlands sampled. For this we need a method that tests the null hypothesis that there is no significant difference in the sample averages (see Chapter 5 for more details). In addition to difference tests between samples, there are also relationship tests between variables, and tests designed to examine associations between categories of variables. Table 1.3 summarises some commonly used statistical approaches to these research questions.

There are various questions that we might ask as part of an investigation, and it is important to be clear about possible analysis methods in advance of any sampling. The choice of test depends not only on the question being asked, but also on the data types being used. Where data are ranked, but not measured (i.e. ordinal data – p. 21), then a suite of tests called non-parametric tests may be used. The alternative (using parametric tests) is more robust and generally preferred, but requires data to be on a measurement scale (i.e. interval/ratio data). Therefore it is usually an advantage to obtain

Table 1.3 Common statistical tests
Note that in each case, there are possible questions (and analyses) dealing with more than 2 samples and/or variables – see Chapter 5 for further details

Example question	Null hypothesis	Type of test	Data required
Is there a difference between the number of birds found in deciduous woodlands and coniferous woodlands?	There is no significant difference between the number of birds in deciduous and coniferous woodlands	Difference tests e.g. a *t* test or a Mann–Whitney U test (see Chapter 5 p. 265)	Two variables: one nominal describing the woodland type and one based on either measurements (i.e. actual numbers) or on a ranked scale that describes the number of birds
Is there a relationship between the number of birds and the size of the woodland?	There is no significant relationship between the number of birds and the size of the woodland	Relationship tests, e.g. correlation analysis (see Chapter 5 p. 269)	Two variables: one (either measured or ranked) that describes the number of birds and one (either measured or ranked) that describes the size of the woodland
Is there an association between whether birds are resident or not and whether the woodlands are deciduous or coniferous?	There is no significant association between the frequency of residency and the frequency of woodland type	Frequency analysis e.g. a chi-square test (see Chapter 5 p. 274)	Two variables: one nominal describing the residency status of the birds and one nominal describing the woodland type

measurement data rather than rankable data wherever possible. Even where measurements are taken, parametric tests may not be the most appropriate. This is because most parametric tests require the data to conform to a type of distribution called a normal distribution. Briefly, this is determined by examining histograms of the data (with the variable of interest plotted on the x axis and the frequency of its occurrence on the y axis) to see whether they have a symmetrical pattern rather than a skewed distribution where the mode (the value with the maximum frequency) lies towards the left or right rather than the centre of the histogram (Figure 1.6). For further details of which test to use, see Chapter 5. There are also different tests depending whether the data are matched or unmatched (see p. 266).

To illustrate some of the considerations in project design and data collection, we start with a research question that sounds relatively simple on the face of it: is there a

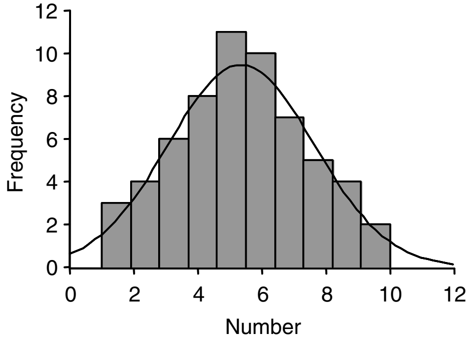


Figure 1.6 Dataset approximating to a normal distribution

relationship between the size of trees and the number of squirrels' dreys in the canopy of the trees? Ideally we would want to measure the canopy height with some degree of accuracy. This would enable us to work out whether the relationship exists using a parametric statistical technique called Pearson's product moment correlation analysis (see Chapter 5). However, it may be difficult even to see the tops of very tall trees and those obscured by other trees. Thus, we may estimate tree height, perhaps into several groupings. We can of course rank these data but this means that we need an alternative approach for analysis that is suitable for ordinal data. This is Spearman's rank correlation coefficient analysis, which is not quite as powerful as the Pearson method. The power of the test is its ability to detect a true relationship (or difference, or association) if one exists. If we knew that any such relationship was likely to be fairly weak, then the less powerful technique might not reveal it and we could be wasting our time in not measuring the trees relatively accurately to obtain measurement data and thus employ the more powerful test. Alternatively, if we are only interested in revealing strong relationships, then using ranked size classes to indicate tree height may be acceptable. We will review these techniques in more detail in Chapter 5. The other complexities in this apparently simple question include ensuring that all other aspects are as constant as possible (e.g. species of tree, surrounding landscape, density of the squirrel colony, etc.).

Predictive analysis

We may wish to collect data to set up a model that enables us to predict the outcome in a hypothetical situation, one of the simplest of which is known as a linear regression model. Thus, if we are interested in looking at a possible relationship between woodland size and the number of birds and knew that this was likely to produce a significant linear relationship, then we may wish to use this fact to calculate the expected number of birds found in any particular woodland. This could be used theoretically or in conservation management to check that we have the sort of bird biodiversity that we expect from other data. Here it is important to note that any such prediction can only be made if the woodland area in which we are interested lies between the minimum and maximum value of the dataset we used to establish the model. We first need to establish which variable is the dependent and which is the independent variable: that is, which is likely to be affected (the dependent variable – plotted on the *y* axis of a scatterplot) by the other (independent variable – plotted on the *x* axis of a scatterplot). Here obviously the number of birds (the dependent variable) is more likely to be dependent on the size of the woodland (the independent variable) than vice versa. To develop such predictive models, data should be measured values. We can extend the technique to cover the case where there are a number of independent variables (e.g. woodland area, habitat diversity, area of associated green space, distance to nearest water body) that might influence the number of birds (see Chapter 5 for further details).

Multivariate analysis

Where the question to be asked is a complicated one involving a number of dependent and/or independent variables then multivariate analyses may be appropriate. The choice of analysis depends on whether the dependent variable is a category, or a ranked or measured variable, and on whether the independent variables are categories, ranked or measured (or even a mixture). Although most (but not all) such analyses only have one dependent variable, there may be multiple independent variables. For example, we may want to know whether the number of birds differs in different types of woodland when we take into account the woodland size (measured variable), woodland type (nominal variable), distance to the nearest neighbouring woodland (measured variable), age of woodland (measured variable) and the land use type surrounding the woodland (nominal variable). Here, we could enter all of the data into one analysis that would take into account the interrelationships between each variable and produce a model describing the relative importance of each variable on the number of birds (this particular example could be analysed using a generalized linear model – see Chapter 5 for further details). Such techniques are powerful but require a full understanding of the data and their attributes and may be quite complex to interpret (see Chapter 5 for further details).

Examining patterns and structure in communities

Ecological datasets can be very complex and difficult to visualise. For example, a dataset might include many variables collected as measurements (including counts), as ranks (e.g. scores of abundance) or in a binary form (e.g. presence or absence data). Chapter 5 introduces a number of techniques for visualizing complex datasets to enable the use of a range of different types of data, although variables with large numbers of zeros (as can occur when surveying relatively rare species), cases where data are heavily skewed, or situations where variables are measured on scales of greatly differing magnitude may require data transformation before using these techniques (see Chapter 5 for further details).

As an example, we might collect information about woodlands on the basis of their size, age, distance to the nearest neighbouring woodland, etc. Since some of these variables will be related to each other, we might wish to find out the underlying pattern of interrelationships within the data and hence identify a number of unrelated factors that can be used instead of our large number of variables. This is a data reduction exercise, reducing the number of variables we have measured into a smaller number of factors that take into account the interrelationships between the variables.

Alternatively, we might wish to look at a range of species found in each of several woodlands and see which woodlands have similar species types. This is a similarity or clustering analysis and depending on the technique used to calculate the similarities, data are normally recorded as a matrix that contains either measurements (e.g. counts), ranks (e.g. ranked abundance) or binary data (e.g. species presence or absence). A similar technique to clustering enables us to visualise patterns in either the individuals (in this example, the woodlands) and/or the variables (here, the types of species). This is known as ordination and there are a number of different methods available depending on the algorithm (i.e. statistical formula). Such methods can utilise data comprising measurements, ranks or binary information.

Choosing sampling methods

The choice of sampling method will usually be dependent upon the habitat type and organisms being studied (see Chapters 2–4). However, all sampling techniques have limitations, and there are some general principles that are applicable to most sampling methods, for example:

- Some techniques may be suitable for a limited range of habitats, or be biased in favour of active rather than sedentary animals, or collect only a subset of the population being examined (e.g. males rather than females, or those migrating rather than those resident). It is therefore very important that limitations are known and dealt with during the design of the research to avoid later problems in interpretation.

- Usually we wish to collect as many data as is feasible, bearing in mind any restrictions in terms of time and personnel. However, different techniques require different skills or time frames, and may collect differing amounts of data, thus our choice may be restricted by logistical considerations.

- Many techniques are not directly comparable with each other, and even using the same technique but under different conditions (e.g. between habitats with very different vegetation layers, between night time and daylight collections, at different times of the year) may not produce comparable data.

- Limitations of the equipment being used may mean that monitoring environmental variables is restricted if, for example, differences between areas are smaller than the accuracy of the equipment allows.
- Resource issues may determine the methods available for use: the cost of equipment, necessity for training, ease of relocation of apparatus between sites, and health and safety issues could all limit the choice of methodology.

Summary

Ensure that you take sufficient time in the planning phase of your research project to cover all of the component parts. This includes health and safety and legal issues as well as making sure that your aims and objectives are focused and that any methods employed are appropriate to gather and analyse data. At each stage, consider the details of the implementation, whether this is in the practicalities of sampling or data management. Box 1.8 gives some general guidelines that should be ticked off in advance of implementing your project.

Box 1.8 Checklist for field research planning

☑

☐ Determine the question and formulate the aims, objectives and hypotheses
☐ Determine whether a manipulative experiment or observational survey would be the most appropriate method to use
☐ Decide whether you are looking primarily at the presence or absence of a species, relative abundance (e.g. counts of organisms), absolute abundance (e.g. population sizes or densities), community structures, behavioural responses, etc.
☐ Determine the statistical analyses that are likely to be employed
☐ Select the appropriate sampling technique, taking into account the intended statistical analysis
☐ Decide on the taxonomic level for identification, or the appropriate ecological grouping you will use (see Chapter 2)
☐ Assess what types of data will be collected and produce a standardised recording sheet
☐ Work out the sampling design including sample sizes
☐ Select appropriate sample sites
☐ Determine the site characteristics that will be monitored and choose appropriate techniques for doing this
☐ Obtain permission to use the sites and check any legal restrictions
☐ Risk assess all the work to be carried out
☐ Employ a pilot study and amend your protocol if necessary