Design and Analysis of Ecological Data
Study Design Concepts:

1. Study design. ................................................................. 2
2. Scope of inference. .......................................................... 3
3. Sampling units............................................................... 7
4. Sources of variability...................................................... 10
5. Sampling designs......................................................... 17
6. Confounding & collinearity. ............................................ 21
7. Replication & sample size. .............................................. 23
8. Randomization.............................................................. 27
9. Independence of units. .................................................. 30

*Much of the material in this section is taken from Quinn and Keough (2002; chapter 7)*
Study Design Concepts...

Study design

*Study design* refers to the number and spatio-temporal distribution of suitably defined sampling units and the particular manipulations (i.e., treatments) and/or observations to be made on each unit for the purposes of answering a specific research question.

The goal of study design is generally to provide the most efficient (in terms of cost) and precise estimates of parameters of the population (suitably defined) given real-world practical and logistical constraints.

1. Study design

Broadly speaking, study design refers to the number and spatio-temporal distribution of suitably defined sampling units (also referred to as experimental and observational units depending on the context) and the particular manipulations (i.e., treatments) and/or observations to be made on each unit for the purposes of answering a specific research question.

The goal of study design is generally to provide the most efficient (in terms of cost) and precise estimates of parameters of the population (suitably defined) given the real-world practical and logistical constraints confronting the study. There are many important factors to consider in achieving this goal.
2. Scope of inference

Study design derives from the research question and is established after the intended scope of inference is determined. Scope of inference refers to the extent over which inferences are to apply, where “extent” is suitably defined as the “population” of interest. Recall that inference involves drawing conclusions about an underlying (statistical) population from a sample and can take on many forms, including estimating parameters of a model representing the system under consideration, testing hypotheses about the underlying system, evaluating competing models of the system, and making predictions about future observations. Thus, scope of inference refers to the “population” we are trying to describe from our sample. Here, “population” is used in the statistical sense (as opposed to the biological) and can take on many meanings depending on the research question. The population should be described in terms of the following three attributes at a minimum:

Spatial
Usually there is a spatial (geographic) dimension to the research question – often defined by the study area. We should always clearly define the spatial extent over which our findings are to apply. This has important implications for selecting sampling units (below) to sample and in choosing the number of units to sample, since one of the important goals of study design is to ensure that the sample adequately represents the “population” for which inferences are to be made. For example, the larger the study area, the greater the number of samples needed to adequately represent the range of variation in the “population”, since larger areas typically are more variable.
Study Design Concepts...

Scope of inference

Scope of inference refers to the extent over which inferences are to apply, where “extent” is suitably defined as the “population” of interest.

- Spatial extent...
- Temporal extent...
- Ecological extent...

Temporally

Usually there is a temporal dimension to the research question, even if it is not explicit in the question. We should always clearly define the temporal extent over which our findings are to apply. Again, this has important implications for study design and interpretation of results. If the question involves a dynamic that occurs over say a 10 year period, then it is likely that the sampling will need to take place over a 10 year or longer period. For example, we cannot examine how climate cycles affect fire disturbance regimes by sampling climate and fires in a single year, since the dynamic under consideration lasts several years or decades (or longer).
Study Design Concepts...

Scope of inference

Scope of inference refers to the extent over which inferences are to apply, where “extent” is suitably defined as the “population” of interest.

- Spatial extent...
- Temporal extent...
- Ecological extent...

Clearly defining the “population” is critical to ensuring a sufficient and representative sample.

Ecological/socio-economic

There will always be an ecological/socio-economic dimension to the scope of inference, where the particular attribute(s) needing attention will depend on the particular question. For example, if my research question involves sampling species A only, then my scope of inference will almost certainly be species A and not species B or C. On the other hand, if my study involves sampling an assemblage of species, say A, B and C, then my scope of inference is likely to be the entire assemblage. And so on. Note, it is quite common for the research question to be much broader than is practical for any single study. In such cases, it is best to define the scope of the overall research program separately from the scope of inference for the study at hand. Given the research question, the ecological scope of inference should be clear and practical, and serve as the basis for the sampling regime.

There may be other dimensions to the question of what constitutes the “population” of interest, but the three listed above are the “big” ones that need to be considered in pretty much every case. Note, it is imperative that we clearly define the “population” before trying to collect a sample from it so that we can ensure that our sample will adequately represent the population. We cannot design an efficient and effective sample if we cannot define the “population” it is suppose to represent. Moreover, after the data are collected and analyzed, we need something to focus our inferences on. Having sample results without a clearly defined “population” to focus them on is like trying score on goal in a soccer game without knowing where the goal is.
Study Design Concepts...
Scope of inference

In practice, it is sometimes useful or necessary to infer beyond the scope of inference

- **Spatial extent**...
- **Temporal extent**...
- **Ecological extent**...

**Inferring beyond the scope of inference**
As described above, scope of inference refers to the spatial, temporal and ecological/socio-economic extent over which our conclusions apply. Thus, the scope of our inference depends on how we defined the “population”, and the strength of our inference depends on the adequacy of our sample (e.g., representativeness, intensity) and statistical model. Technically, we cannot infer our findings beyond the scope of the “population” sampled. However, in practice, it is often useful to extend our findings beyond the formal scope of inference. For example, let’s say we studied a local population of an endangered species and documented significant habitat relationships. Technically, our findings apply only to the one population we studied. However, we would be remiss in our ethical obligation to conserve the species if we did not extend our findings to the broader region when called upon to develop a recovery plan for the species. Thus, we should not let statistics get in the way with making good use of our findings. We always have to balance the limitations of formal statistical inference with the practical real-world need to extend our findings as far possible.
Study Design Concepts...

Sampling units

The “population” is the collection of all possible sampling units

- **Grain...** units represent the “grain” or lower limit of resolution of the study, and may be natural or artificial units

3. Sampling units

A corollary to defining the statistical population – and the scope of inference – is the delineation of sampling units (also referred to as “experimental units” when used in an experimental study and “observational units” when used in an observational study, see below), since the “population” is the collection of possible such units. Choosing the right sampling unit to match the research question is critical and not as trivial as it might seem. There are several factors to consider:

*Sampling units represent the grain or resolution of study*

Whereas the scope of inference deals with the extent or domain of the study and the statistical population, the sampling unit deals with the grain or resolution of the study and the particular entities that comprise the population. As such, the sampling unit is the lower limit of resolution in our ability to make inferences from the model. As with the population, the sampling units must be clearly defined in terms of their spatial, temporal, and ecological/socio-economic attributes, which must match the scale of the research question perfectly. Sampling units may be natural units (e.g., individuals, ponds, homes) or artificial units (e.g., plots, quadrats)
Study Design Concepts...
Sampling units

The “population” is the collection of all possible sampling units

- Free to vary... units are always the finest unit that is free to vary in both independent and dependent variables

Q: Do streets with trees confer greater home energy efficiency?

Subsamples?

Sampling units are free to vary
The sampling units are always free to vary; i.e., each unit has the potential to take on a different value of the variable(s) of interest. This is key to determining what constitutes the right unit of observation. Generally, the units must have the potential to take on different values of both the independent (explanatory) and dependent (response) variables in order to be deemed the units of observation in the model. For example, let’s say we are interested in how the removal of street host trees for the Asian longhorn beetle in Worcester Massachusetts effects home energy efficiency (by increasing exposure to wind and sun). Home energy efficiency is measured at the scale of the individual homes - because it is free to vary among homes – so it is tempting to think of the homes as the observational units. However, the independent variable, street trees, varies at the scale of streets, not individual homes, because entire streets are treated the same. In this case, the appropriate sampling units, given the question, are the streets. Accordingly, we might sample a single home on each street and assume that it represents home energy efficiency for the street or, better yet, sample several homes on each street and compute the average energy efficiency of homes on each street. To treat the home as the sampling unit in the model would be a form a pseudoreplication (sensu Hurlbert 1984). For example, if we measure energy efficiency for 12 homes distributed among 4 streets, 2 treated and 2 controls, we really have only 4 independent observations. If we treat each home as an independent observation, we would triple our degrees of freedom and thus make any hypothesis test appear more significant than it really is. In this case, the home is referred to as a subsample; it is the actual unit in which we record the dependent response (energy efficiency), but it is below the resolution at which the independent variable(s) is measured/observed (street trees).
Study design concepts

Sampling units

The “population” is the collection of all possible sampling units

- **Multiple levels... units can exist at multiple levels, e.g., when they occur in a nested hierarchy of scales**

---

**Multi-level sampling units**

In many cases we have more than one level (or scale) of sampling unit. For example, it is quite common to have units of variation nested with broader units of variation. It is easiest to understand this concept by example. Let’s say we are interested in environmental determinants of reproductive success in spadefoot toads, a seasonal pond-breeding species in the Provincelands of Cape Cod. We might hypothesize that environmental factors operate at multiple spatial (and temporal) scales to influence spadefoot reproductive success. There might be local, within-pond habitat features that operate very locally to determine successful breeding at the “plot” level; e.g., shrub cover and water depth. However, there might also be pond-level features, such as pond size and hydroperiod, and broader landscape-level features, such as proximity to other ponds and upland vegetation conditions, that influence local breeding success. Thus, we might hypothesize that local breeding success is a function of local plot-level habitat conditions, conditioned on suitable pond-level conditions, conditioned on suitable landscape-level conditions. In this case, our question involves evaluating reproductive success at the plot level – our basic unit of interest in terms of the response variable, but incorporating habitat explanatory variables that vary at the plot, pond and landscape scales. We have two choices. First, we can treat each scale separately by treating measurements below the focal scale as subsamples and conducting a separate analysis for each scale. Second, we can incorporate all three levels of habitat variation into the same model using a “mult-level” modeling approach – which we have not addressed yet in this course. The first approach is conceptually and technically quite simple, but the later approach is considered more powerful.
4. Sources of variability

Once the “population” and sampling units are clearly defined (based on the research question), study design focuses on dealing in appropriate ways with the important sources of variability effecting the phenomenon under consideration. One of the critical steps in study design is to identify all potentially important sources of variability and develop a strategy for dealing with each. It may be useful here to distinguish between experimental studies and observational studies, since these are essentially differing strategies for dealing with sources of variability.

1. Manipulative experiment
Manipulative experiments involve a priori specification of one or more treatments and the imposition of the treatment(s) by the researcher as part of the study. The focus in such studies is on determining cause and effect relationships between the treatment(s) and the response. Here, there is a high degree of control over the sources of variability. For example, in the home energy efficiency example, we might remove street trees from randomly selected streets and compare home energy use between treated streets and control streets without tree removal. Typically, all sources of variability are “controlled” for by allocating treatments to experimental units randomly, which is a way of ensuring that sources of variability are not systematically confounded (indistinguishable) with the treatment - more on this below. The only systematic difference between experimental units, therefore, is the treatment(s). There are many standard designs for conducting manipulative experiments, which are often described in association with the method known as analysis of variance, since these designs typically focus on partitioning the variance between the treatments and the inherent variability among experimental units.
2. *Comparative mensurative experiment*... treatments are natural or pre-existing (not imposed); focus is on causal effects of “treatments”

Moderate degree of control over sources of variability

Study Design Concepts...
Sources of variability

Study design is about dealing in appropriate ways with important sources of variation effecting the phenomenon under consideration.

- *Comparative mensurative experiment*... treatments are natural or pre-existing (not imposed); focus is on causal effects of “treatments”

Treatments: late-seral forest area & configuration

2. **Comparative mensurative experiment**

Comparative mensurative experiments involve a priori specification of one or more treatments and are similar in all respects to manipulative experiments except that the treatments are not imposed by the researcher. Instead, experimental units are selected from natural or pre-existing “treatment” groups. The focus in such studies is on causal relationships, but because there are many uncontrolled factors in a comparative mensurative study, the degree of causation that can be inferred is relatively weak. Indeed, it may be impossible to infer true causation at all from a comparative mensurative study, because it is impossible to show that the differences between treatment groups are solely due to the treatment. For example, in my study of breeding bird community response to late-successional forest loss and fragmentation, I used a comparative mensurative design in which late-seral forest area and configuration were the treatments, but I did not impose them. Instead, I randomly selected experimental units (subbasins) that pre-existed in levels of each treatment factor. The treatments were imposed, but they were imposed years before by different entities without a common eye towards my scientific study. Thus, there is no assurance that the units in each treatment level (e.g., 20% forested and highly fragmented) are not confounded with other important sources of variation affecting breeding bird communities. Despite the limitations, comparative mensurative experiments are far more common than manipulative experiments in field ecology due to the difficulty (and cost) of implementing the latter.
Study Design Concepts...
Sources of variability

Study design is about dealing in appropriate ways with important sources of variation effecting the phenomenon under consideration

- Observational study... no treatments per se, differences among sampling units are measured or observed; focus is on association

Low degree of control over sources of variability

3. Observational study
Observational studies involve no treatments per se; that is, sampling units are not selected to provide explicit control over one or more “treatment” variables, for example as was the case in the comparative mensurative experiment. The differences among sampling units with respect to the sources of variability is determined post-hoc after collecting measurements from the sampling units. The focus in an observational study is on association; i.e., determining how the response variable varies in relation to the explanatory variables. However, typically there is no a priori control over the range of variation sampled – it is outcome rather than an input. For example, in the spadefoot toad example, the sampling units were selected by systematically sampling plots at fixed intervals along regularly spaced transects, with a random start to the first transect to introduce some level of randomization to the selection of sampling units. Within each sampling unit each of the explanatory variables (habitat variables) and the response variable (larval abundance) were measured. In this case, the goal is to associate larval abundance with the various habitat factors.
In addition to these basic study design types, strategies for dealing with sources of variability generally fall into four categories depending on whether the source of variability is explicitly controlled or uncontrolled:

1. **Eliminate source of variability**

   One approach for dealing with sources of variability is to eliminate them by design; specifically, limit the “population” to sampling units that exhibit a single or narrow range of variability with respect to that source. For example, if we are interested in the effect of street trees on home energy efficiency, but we recognize that the age of the home, dimensions, architectural style, and building materials all are likely to effect home energy efficiency, then we could limit our study to homes of a certain age, dimension, architectural style and building material to effectively eliminate these as variables. The consequence of eliminating sources of variability in this manner is to reduce the scope of inference of the study, so this should be done only with a clear understanding of the tradeoffs. This approach is the basis for much of experimental science, in which sources of variability are carefully controlled and only the one or two treatment variables are allowed to vary. “Common garden” experiments are an example of this type of study, where the environment is held in common among all experimental units except the one or two treatment differences that are imposed. While this has common application in laboratory studies, it is much more difficult to implement in a field study.
2. “Fix” source of variability...

Another approach for dealing with sources of variability is to “fix” the levels of variability by design; specifically, limit the “population” to units within a priori specified levels of variability of the variable. This can be done via experimental manipulation, for example by imposing different treatments, or by comparative mensuration, for example by selecting and measuring units that naturally exhibit the differences sought. For example, if we were interested in examining how the loss and fragmentation of late-successional forest effects breeding bird communities, we could select landscape units, impose treatments to reduce and fragment the forest, and then measure the bird community response over time in relation to treatment, of course while also measuring the bird community over time in untreated (control) sites. Alternatively, we could select existing landscape units that vary in extent and configuration of late-successional forest and compare the bird community among units in relation to existing “treatment” levels. Here, we have substituted space for time to make this study feasible within a few year period; otherwise, to follow through with the manipulative experiment could take decades. Note, in this strategy we are fixing the levels of variability for the variable or variables of interest; i.e., we are not allowing them to freely vary over their full range of variability. Consequently, our scope of inferences is reduced to the levels of variability investigated. As before, this should be done only with a clear understanding of the tradeoffs between control over sources of variability and scope of inference.
3. Model source of variability post-hoc... ignored in sampling design but accounted for in the model

What is the consequence?

Shrub cover and water depth?

3. Model source of variability post-hoc
Another approach for dealing with sources of variability is to simply measure the source of variability as part of the sampling regime – without a prior control – and simply deal with the influence in the analysis. In this approach, we give explicit recognition to the source of variability but make no attempt to explicitly control it in the study design. Instead, we simply try to account for its effect in the model by modeling it as a random effect on the outcome. For example, in the spadefoot toad example, we recognized that shrub cover and water depth at the plot level are potentially important sources of variation affecting reproductive success, but we did not control for these sources of variability in the sampling design. We simply measured these variables at each plot as they happened to occur. On the positive side, this approach allows us to deal with sources of variability for which a prior control would be exceedingly difficult – which includes many factors in most ecological studies. On the negative side, this approach risks sampling only a limited range of the variability or biasing the sampling towards certain levels that are prevalent (which may be a good thing), and risks confounding among sources of variability (i.e., inability to distinguish between two or more variables that covary).
Study Design Concepts...
Sources of variability

Strategies for dealing with sources of variability in experimental and observational study designs

4. Ignore source of variability...
   ignore in both sampling
design and analysis;
subsume in model error

What is the consequence?

Genetic diversity (heterozygosity)
of breeding adults?

4. Ignore source of variability
A final approach for dealing with sources of variability is to simply ignore them entirely in both the sampling design and the analysis. In this approach, we recognize the potential source of variability but make no attempt to address it in the model, typically because it is too difficult or costly to measure. Here, we let the source of variation be subsumed in the model error; i.e., the unexplained variation in the model. Of course, good planning should avoid the need to adopt this strategy too often, but there will be times that this approach is adopted for practical reasons. For example, in the spadefoot toad example, we might recognize that genetic diversity of the breeding individuals at a site is an important determinant of reproductive output, but realize that it is too difficult or costly to collect the necessary data to include it in the model. It is important to recognize the tradeoffs. The goal in modeling is to minimize the unexplained error, because this leads to powerful inferences. The more sources of variability that are subsumed in the error, the weaker the model and resulting inferences will be.
5. Sampling designs

Study design is ultimately about determining where and when to sample, and there are some common sampling strategies to consider:

1. Simple random sampling

Simple random sampling is where all the possible sampling units in our “population” have an equal chance of being selected in a sample. Technically, random sampling should be done by giving all possible sampling units a number and then choosing which units are included in the sample using a random selection of numbers. In practice, however, especially in field ecology, this method is usually too difficult because the sampling units do not represent distinct natural units or it is not practical to enumerate all of them in the “population” in advance, and a more haphazard approach is used instead where the units are chosen in a less formal but nonetheless random manner. For example, if we wanted to determine the relationship between vernal pool wood frog occupancy and the percent of surrounding uplands forested, we could take the map of all potential vernal pools in an area and pick a random sample of pools to survey for wood frog occurrence and upland forest extent. Alternatively, if we did not have a comprehensive inventory of all vernal pools, we could instead pick locations at random and then search for the nearest pool to survey. You almost can’t go wrong with simple random sampling. However, it may be less efficient than other methods when there is identified heterogeneity in the population or we wish to estimate parameters at a range of scales.
2. *Stratified sampling*... where the “population” is divided into strata that represent clearly defined groups of units within the “population” and we sample independently (and randomly) from each of these groups. For example, we may wish to estimate characteristics of a “population” of forested wetlands in an area. Let’s say our variable of interest is species richness of aquatic macroinvertebrates. If the forested wetlands clearly fall into different wetland types, e.g., coniferous, deciduous, and mixed, then we might take random samples of wetlands from each wetland stratum separately. Stratified sampling is likely to be more representative in this case than a simple random sample because it ensures that the major wetland types are included in the sample. Usually, the number of units sampled from each stratum is proportional to the total number of possible units in each stratum or the total size of each stratum (e.g., area). Estimates of population parameters are adjusted accordingly by including the strata as a predictor variable in the model.
Study Design Concepts...

Sampling designs

Strategies for where and when to sample in experimental and observational study designs

3. Cluster sampling... especially designed for situations where the sampling units are naturally organized in a hierarchy

Salmon within watershed (1st), within reach (2nd), within channel unit (3rd)

3. Cluster (hierarchical) sampling
Cluster sampling also uses heterogeneity of the population to modify the basic random sampling design. Cluster sampling is especially designed for situations where the sampling units are naturally organized in a hierarchy. For example, say we can identify primary sampling units (clusters) in a “population”, e.g., individual watersheds. For each primary unit (watershed), we then record all secondary units, e.g., stream reaches, within each watershed. Simple cluster sampling is where we record all secondary units within each primary unit. Two stage cluster sampling is where we take a random sample of secondary units within each primary unit. Three stage cluster sampling is where we take a random sample of tertiary units (e.g., channel units) within each secondary unit (e.g., reach) within each primary unit (e.g., watershed). Simple random sampling is usually applied at each stage, although proportional sampling can also be used in which we sample the same proportion of units within each higher level unit. As before, estimates of population parameters are adjusted accordingly by including the strata as predictors in the model.
Study Design Concepts...
Sampling designs

Strategies for where and when to sample in experimental and observational study designs

4. **Systematic sampling**... where we choose sampling units that are equally spaced, either spatially or temporally.

Systematic sampling is where we choose sampling units that are equally spaced, either spatially or temporally. For example, we might choose plots along a transect at fixed intervals or we might choose regularly occurring sampling dates. In the spadefoot toad study, for example, we sampled plots every 10 m along each transect and sampled them every three weeks during the breeding season. Systematic sampling is sometimes used when we wish to describe an environmental gradient and we want to know where changes in the environment occur. For example, let’s say we want to measure ponderosa pine establishment away from the nearest mature forest edge (i.e., seed source). Simple random sampling away from the source might miss the crucial region where the seedling establishment undergoes rapid change. Sampling at regular intervals is probably a better choice. The big risk with systematic sampling is that the regular spacing may coincide with an unknown environmental gradient and so any inference to the whole population of possible sampling units would be biased. Systematic sampling can have a single random starting point, where the first unit is chosen randomly and then the remainder evenly spaced. Alternatively, a cluster design could be used, where clusters are chosen at random and then systematic selection on secondary sampling units within each cluster is used. For example, we might select watersheds at random and then sample every 10th stream reach upstream from the outlet within each watershed.
6. Confounding & collinearity

An important issue in study design is that of confounded sources of variation and the closely allied concept of (multi)collinearity.

Confounding
Confounding means that the effects of two (or more) independent variables cannot be isolated from each other. Confounding is often used in reference to experimental designs where a treatment variable is confounded with another (potentially unmeasured) factor that might be causing the observed response. For example, let's take the home energy efficiency example. Let's say our design involves sampling houses on streets with and without street trees, but that all the streets we sample with street trees happen to be on north aspects while all of the streets we sample without street trees happen to be on south aspects. If we observe a difference in energy efficiency between streets with and without street trees, we cannot say it is due to the differences in street trees, since the sampling units also differ in aspect. The “treatment” variable (street trees) is completely confounded with the variable aspect, which we might reasonably expect to influence home energy efficiency due to differences in solar radiation on north and south aspects.
Collinearity refers to correlated explanatory variables and is often used in the context of multiple regression, where two or more explanatory variables are correlated with each other and the observed response cannot be unequivocally attributed to the individual explanatory variables. Multicollinearity refers to correlations involving more than two variables. For example, in the spadefoot toad example, if shrub cover and water depth are strongly correlated at the plot level, then it will be impossible to distinguish their influence on reproductive success. In the context of multiple regression, collinearity results in unstable parameter estimates, meaning that small changes in the data or deleting one of the explanatory variables can change the estimated regression coefficients considerably, even changing their sign. In addition, standard errors of the estimated coefficients are inflated when some of the explanatory variables are correlated. Therefore, the overall regression might be significant when none of the individual regression coefficients are significant.

Approaches for dealing with confounding and collinearity vary. All approaches begin by being aware of all potential sources of variability. In the simplest situations, each source of variability might be explicitly controlled via study design, thus avoiding strong confounding. However, in more complex situations, it may not be possible to avoid the potential for confounding and the inferences must be adjusted accordingly. For example, instead of concluding that the observed differences are due to ‘x’, it may be necessary to say they are due to “x or y”. In the multiple regression context, there are more sophisticated methods for dealing with collinearity among variables in the model, but they go beyond the scope of our discussion.
7. Replication & sample size

One of the key principles of study design is the necessity of replication, both in the context of an experiment and in an observational study.

Replication

Replication is usually used in the context of an experiment, where treatments are applied either directly (through experimental manipulation) or indirectly (through observation of existing treatment differences), the former being referred to as a manipulative experiment and latter as a comparative mensurative experiment (sensu Hurlbert 1984). In the experimental context, replication means having replicate observations (sampling units) at a spatial and temporal scale that matches the application (or observation) of the experimental treatments. Replicates are essential because ecological systems are inherently variable. Linear model analysis of designed experiments usually rely on comparing the variation between treatment groups to the inherent variability among experimental units within each group. An estimate of this latter variability requires replicate units. Replication at an appropriate scale also helps to avoid confounding treatment differences with other systematic differences between experimental units. For example, if we sampled homes on a single street with street trees and homes on a single street without street trees, then the effects of our treatment (street trees) is completely confounded with inherent differences between the two streets related to their spatial location, such as slope. With two or more replicate streets for each of the two treatments (with and without street trees), we can be much more confident in attributing differences
between treatments to street trees rather than inherent street differences. Note that replication does not guarantee protection from confounding because it is still possible that, by chance, all our streets with street trees are different from our streets without street trees in some other way that influences home energy efficiency. However, the risk of confounding is reduced by replication, especially when combined with randomized allocation of treatments to experimental units (below).

While most ecologists are well aware of the need for replication, we often mismatch the scale of those replicates relative to treatments being applied. For example, in the home energy efficiency example, it would be easy to mistake the sampling of multiple homes on a single street for replication. Recall that the treatment is applied at the level of the entire street, not individual homes. So sampling multiple homes, while necessary, is not the same as sampling multiple streets. Sampling multiple homes on a single street is a form of “subsampling”, and would be highly desirable if there is considerable inherent variation in energy efficiency among homes on the same street.

In extreme cases, it may be difficult or impossible to obtain replication at the appropriate scale. There are a number of options in such cases, although none of them are particularly ideal. One possibility is to run the experiment a number of times, each time switching the treatments between experimental units. In cases involving the impacts of human activities, there may be only one place at which the potential impact occurs; e.g., toxic discharge site. Special designs for handling such cases have been proposed, but they are not without substantial limitations and criticism.
Study Design Concepts...

Repetition & sample size

Ability to evaluate treatment effects and estimate population parameters precisely

- **Sample size**... our ability to estimate population parameters precisely is directly related to the number of observations we have

- Our ability to sufficiently describe ecological relationships is directly related to sample size

**Sample size**

The need for replication is closely allied to the need for an adequate sample size in an observational study. Whereas replication deals with the need for replicate units within treatment levels, in an observational study without treatments per se there is an equivalent need for multiple samples to obtain a reasonable representation of the “population”. It almost goes without saying that our ability to estimate population parameters precisely is directly related to the number of observations we have. Recall that we already discussed sample size in the context of statistical power in the chapter on stochastic simulation. The examples presented there demonstrated that statistical power, our ability to correctly reject the null hypothesis, is directly related to the sample size. More generally, it can be shown that the precision in our estimates of population parameters is directly related to sample size as well. This is because the precision in parameter estimates is usually given by the standard error of the estimate and standard errors decrease with increasing sample size. Recall that the standard error of the mean is equal to the standard deviation divided by the square root of the sample size.

Aside from the issue of statistical power (in the general sense), sample size is critical to the issue of **sufficiency**. Sufficiency has to do with whether an ecological relationship is sampled enough to reliably characterize it, and it is perhaps more an ecological consideration than a statistical one. For example, in the spadefoot toad example, let’s say our sample of 100 plots included only 1 plot with 100% shrub cover. The relationship between shrub cover and reproductive success at the extreme end of the shrub cover gradient is not sufficiently sampled to be able to infer anything ecologically reliable
about the relationship at that point on the gradient. Increasing sample size is one way to improve the likelihood that each ecological condition is sufficiently represented in the sample. The other way is to use a stratified or systematic sampling design to ensure equal and sufficient sampling of the full range of the explanatory variable.
Study Design Concepts...

Randomization

Ability to evaluate treatment effects and estimate population parameters in an unbiased manner

- Random sampling... our ability to estimate population parameters without bias requires some level of randomization to the selection of sampling units

- Key to identifying the true scope of inference is identifying where the randomization is applied

8. Randomization

One of the key principles of study design is the necessity of randomization in the selection of sampling units to ensure unbiased estimates of treatment effects and of population parameters.

Random sampling

Unbiased sampling is necessary in order for inferences about the “population” to be valid. Randomization is the mechanism by which we achieve this goal. Every study design should involve some level of randomization in the selection of sampling units. We already partially addressed how we might achieve randomization using different sampling designs: simple random sampling, stratified sampling, cluster sampling and systematic sampling. In practice, however, there are myriad modifications to these and other sampling designs needed to accommodate the many and varied real-world conditions encountered in field ecology, but the principle of randomization is always maintained - or should be. A key to identifying the true scope of inference of a study is identifying where the randomization was applied. For example, in the home energy efficiency example, if streets are selected at random from the “population” of streets in the city, then the inferences will apply to the city. If on the other hand, streets are selected at random from a particular neighborhood, because that is where we happen to be working, then the inferences will apply to that neighborhood and not the city.
Study Design Concepts...

Randomization

Ability to evaluate treatment effects and estimate population parameters in an unbiased manner

- Random allocation of treatments...
in an experimental study, random allocation of treatments to experiment units is the usual mechanism for achieving unbiased estimates of treatment differences while also avoiding the potential for confounding

Random allocation of treatments
One of the goals in the design of an experiment is to assign treatments to experimental units in such a manner as to avoid the potential for any systematic bias. Random allocation of treatments to units is the usual mechanism for achieving this goal. In a manipulative experiment involving the imposition of treatments to units, the usual approach is to randomly assign treatments to units. In a comparative mensurative experiment involving the comparison of naturally differing experimental units, the usual approach is to randomly select units from each pre-existing treatment type. In both cases, it means that no pattern of treatments across experimental units is subjectively included or excluded and that systematic differences between experimental units that might confound our interpretation of treatment effects are minimized. For example, in the home energy efficiency example, a manipulative experimental design to ensure unbiased results might be to identify the “population” of suitable streets and then randomly assign streets to the treatment group (remove street trees). This should ensure that any other factors influencing home energy consumption is not systematically confounded with the treatment. Of course, this will likely depend on the number of replicates, since with too few replicates it is quite possible that some confounding will occur by chance. In a comparative mensurative design, we might identify all streets currently with and without street trees and randomly select some for inclusion in each group. Of course, we would need to consider whether the two pre-existing groups are biased in any meaningful way that might confound our treatment results with some other pre-existing difference between groups.
Study Design Concepts...

Randomization

Ability to evaluate treatment effects and estimate population parameters in an unbiased manner

- Randomization vs interspersion...
  treatments should also be well interspersed (mixed together) in space and/or time to ensure that treatments are not confounded with some spatial or temporal environmental gradient.

Randomization versus interspersion
The allocation of experimental units to treatments raises the difficult issue of randomization versus interspersion. The idea behind interspersion is that treatments should be well interspersed - mixed together - in space and/or time to ensure that treatments are not confounded with some spatial or temporal environmental gradient. For example, in the home energy efficiency example, if the study area exists on an elevational gradient, we wouldn’t want our treatment groups to be confounded with elevation by chance; e.g., all street units with trees at lower elevations and all street units without trees at higher elevation. Interspersion of treatments ensures that units in both groups are found at all elevations. Interspersion of treatments can cause problems of its own; for example, if treatments are interspersed in a regular fashion, alternating treatment A and B to units along a gradient, then there is a chance of confounding with some periodicity in other environmental variable. The issue of randomization versus interspersion illustrates one of the difficult challenges of experimental design.
9. Independence of units

Another important issue in study design is that of independence of sampling units. The lack of independence among sampling units will make interpretation of results difficult and may invalidate some forms of statistical analysis. Recall that in our discussion of likelihood we considered the likelihood of the entire data set by taking the product of the likelihoods of individual data points (or the sum of the log-likelihoods). Joint probability only works in this manner when the observations are independent. Independence of observations is a huge concern in ecological studies, because in ecological systems where things are interconnected it is difficult to achieve complete independence, which is assumed in most statistical analyses.

**Autocorrelation**

Much of the attention has been on spatial (and temporal) dependencies among sampling units, since the strength of ecological interactions generally decreases with distance (and time). Things closer together (in space and time) are more likely to behave similarly than things farther apart. This is generally known as “autocorrelation”. The concept is quite simple. Consider an example in which we would like to know whether tree growth is affected by soil moisture. If we sample a tree and the corresponding soil and then move 10 m away and sample another tree and its soil, we are likely to find a very similar set of values for tree growth and soil moisture because they are on the same soil facet. If we move 1,000 m and sample another tree and its soil we are likely to be on an entirely different soil facet. It is equally likely that the new point will be similar to the first point or
completely different in growth and moisture because it is an entirely different soil facet and will depend on whether by chance the point happens to fall in a soil type with similar moisture conditions. In this situation we would find that the soil moisture conditions (and presumably tree growth) are correlated up to some distance; that is, there is a spatial dependency in soil moisture conditions up to the scale of the dominant soil facet.

The same sort of thing happens with temporal data, where sample points close together in time exhibit a positive autocorrelation. For example, consider the analysis of river discharge. If we sample discharge on consecutive days, we will find that the discharge on one day can be predicted pretty well by the discharge on the previous day. However, if we get far enough apart in time, say several months, the discharges will be unrelated. Another common source of lack of independence is when we repeatedly the same units over time. For example, let’s say we are interested in the relationship between river discharge and percent imperviousness of the watershed and we had data for several watersheds over several decades. We would likely find that the repeated observations in a single watershed are more similar to one another than the observations from other watersheds because each watershed has unique hydrological properties. To ignore the grouping of observations by watershed (i.e., the lack of independence of observations) would be a form of pseudoreplication (sensu Hurlbert 1984).
Study Design Concepts...

Independence

Strategies for dealing with independence issues

1. *Ensure by design...* select sampling units at a minimum distance (in space and/or time) based on ecological knowledge

Tradeoff between independence and randomization?

Strategies

The consequence of treating observations as independent when in fact they are not, for example because of spatial or temporal autocorrelation, is generally an underestimation of the true residual variance (i.e., model error) and seriously inflated Type 1 error rates for hypothesis tests on parameters. Strategies for dealing with lack of independence loosely fall into three categories:

1. *Ensure independence by study design.*—most study designs incorporate independence into the selection of sampling units; that is, sampling units are selected a priori in a manner that strives to achieve statistical and biological independence among units. For example, many radio telemetry studies of organisms involve getting periodic locations after an interval that allows the individual the opportunity to access their entire home range. In addition, it is quite common to enforce a minimum distance among sampling units based on some prior ecological knowledge of the system. For example, choosing bird survey points that are farther apart than the average home range diameter. However, rarely, if ever, is it possible to ensure independence by design, since we rarely have sufficient a priori knowledge to do so. More importantly, by enforcing this criterion or any other independence criterion, the sample is no longer truly random, so there is a tradeoff between independence and randomization.
Study Design Concepts...

Independence

Strategies for dealing with independence issues

2. Model dependencies... techniques exist for examining independence of observations and explicitly incorporating any dependences into the model

Sometimes autocorrelation is the focus of the study, e.g., time series analysis

2. Model dependencies.—since we can never ensure true independence, it is usually a wise idea to evaluate the lack of independence post-hoc; i.e., after collecting the data. This can be as simple as examining the correlation between the residuals of a model and the distance (in space or time) between observations to determine the magnitude of the problem. These sorts of autocorrelation analyses are quite common and can be used to confirm that autocorrelation is not a serious problem. Sometimes these results are used to advise subsequent sampling, for example, to identify the distance (in space or time) at which observations become statistically independent. If the autocorrelation is serious, either by the intent of the study design or by an accident of nature, there are a variety of procedures for modeling the autocorrelation explicitly; that is, incorporating into the model a means for handling the lack of independence. These procedures can be fairly sophisticated and go well beyond the scope of this chapter, so at this point simply be aware that procedures exist for dealing with non-independent data. It should also be noted that in some cases the focus of the study is modeling the autocorrelation itself. Time series analysis is a suite of procedures devoted to analyzing data that is serially autocorrelated and the focus is to describe the patterns of autocorrelation. In this context, lack of independence is not a problem; quite the contrary, it is the pattern of interest.
Study Design Concepts...

Independence

Strategies for dealing with independence issues

3. *Ignore the problem...* the degree to which this is deemed “acceptable” will largely depend on the research question, study design and type of inferences made.

Lack of independence is particularly troublesome in the context of hypothesis testing.

3. *Ignore the problem.*—this is not really an acceptable strategy, but it is mentioned here because in practice investigators often simply ignore the independence issue altogether. The degree to which this is deemed “acceptable” will largely depend on the research question, study design and type of inferences made. For example, independence is particularly troublesome in the context of conventional hypothesis testing. If hypothesis testing is of no concern, then lack of independence may not be a major concern. Moreover, if the sample is drawn from a population in a way that reflects the inherent dependence among sampling units in a population, then the estimates of population parameters and the precision of those estimates may apply perfectly well to that population.