Replication

Given the importance of replication to Power and statistical inference it is important that we spend some time thinking about what true replication is and the importance of independence of samples.

Replication

Statistics is based on collecting independent samples from a population. As we discussed in the Sampling and Power handout the *Law of Large Numbers* states that the degree to which those samples reflect the true population depends on the number of samples collected. It is essential that samples are independent. Hurlbert (1984) defines independence to have been achieved when pairs of samples treated alike are on average no more similar or dissimilar than those treated differently. Another way of saying this is that samples take on their unique values as a result of separate causes than have led to values that are unique from other points.

If you were a freshman student who was just learning the importance of sample size to statistics you might think its a good idea to copy and paste your data 10 times before analyzing them so that you would have 10x the sample size and much more power!!!! It would first be unethical to fabricate data, but it is also inappropriate statistically because each of those data points in exactly the same as 9 others in the dataset because of the same underlying biological causes. They are not independent!

Pseudoreplication

Pseudoreplication involves drawing inferences based on replicates that are not truly independent. That is, they are not true replicates. They are "pseudoreplicates". This is a problem in statistics because of the importance of sample size. So in some ways it is an issue of using an inappropriate df to test your hypothesis. **The problem is that pseudoreplication makes your statistical test too liberal** (increases your Type I error rate).


Ecological Monographs, 54(2), 1984, pp. 187–211
© 1984 by the Ecological Society of America

PSEUDOREPLICATION AND THE DESIGN OF ECOLOGICAL FIELD EXPERIMENTS[1]

STUART H. HURLBERT
Department of Biology, San Diego State University, San Diego, California 92182 USA

Pseudoreplication is not just a problem of experimental design. It is a combination of inappropriate design and statistics for a particular hypothesis. Hurlbert (1984) brought this issue to

light in his classic paper in *Ecological Monographs*. This is a citation classic that has been cited over 8800 times! That is a just about a citation every second day! It is also worth noting that Hurlbert is widely regarded as an intolerant and racist person. Citation counts are a dubious measure of scientific achievement and an even worse measure of personal character!

In my experience folks in Ecology and Evolution are also particularly prone to commit pseudoreplication. In an informal sample of 31 manuscripts that I reviewed over a 3-year period, 8 of them (26%) had some form of pseudoreplication and were rejected as a result.

Mensurative vs. Manipulative Experiments

Experiments can be described as mensurative or manipulative. Mensurative experiments involve collecting samples in space or time. They sometimes involve elaborate procedures to collect the relevant data that can sometimes seem like a manipulation but they are comparative or observational by nature. For example, you might perform behavioural trials by recording a squirrel's behaviour in an arena or measure lignin content of a moss using some fancy lab chemistry, but the measurements are fundamentally observational. In this case ***be sure to replicate the things you want to make inferences about.***

Manipulative experiments involve two or more treatments, which are applied to experimental units of interest. These are what we think of as true experiments. In this case ***manipulate and replicate the thing you want to make inferences about***. Also, if you want to generalize the effects of the manipulation then you need to replicate the manipulation as well.

Remember that while we might use organisms as tools to measure a process, they are not necessarily the things we want to draw inferences about. Ask yourself:

Are the organisms or the process of primary interest?

Are the organisms just used to measure the process?

If so, then be sure to replicate the process and not just the organism (i.e. the tool).

For example, we might measure the effects of competition by a change in plant height. However,

Some Useful Terminology

Experimental Unit (EU) - the smallest unit of replication, in space, time or organisms to which treatments are applied independently and in replication

Treatment - a categorization (mensurative) or manipulation of EU's.

Response - Dependent variable measured at level of EU. If measured at level below EU then this sub-sampling must be accounted for.

competition is a *process* that is characteristic of a population or community. So if you are interested in testing an effect of competition then you need to replicate the process (population or community) and not just the tool used to measure that process (the plant). If you were testing a hypothesis about competition using individual plants as your replicates then this would be pseudoreplication.

When should treatments be replicated?

If your hypothesis refers to a single specific effect or manipulation then you do not need to replicate this manipulation. However, if you want to generalize your conclusions then you need to replicate the manipulation. For example, studies of the effects of reintroducing wolves to Yellowstone National Park do not have any treatment replication. It is not needed because the inferences are to be drawn about wolves in Yellowstone only. These studies, however, cannot conclude anything about wolf reintroductions in general because there is no treatment replication. There might be similar considerations with respect to effects of environmental contamination events,

Types of Pseudoreplication

1. **Simple** - Single replicate per treatment and multiple measures taken of the same replicate (e.g. whole lake experiments).

2. **Temporal** - Similar to Simple but multiple samples taken through time are treated as independent.

3. **Sacrificial** - Studies that start out with true replication but pool samples across replicates when found to not differ.

Examples of Pseudoreplication

Treatments can't be applied or aren't applied at the appropriate level. For example, a hypothesis about the effects of light levels on plant growth is fundamentally about plants. Plants should be the

experimental unit to which treatments are applied. However, it is often too costly to have each plant assigned to its own incubator. Instead we typically put a bunch of plants into the same incubator. In this case the light level treatment is applied to groups of plants in an incubator rather than individual plants. So observed values of growth for individual plants from the same incubator are not independent. So they are not true replicates, but are instead pseudoreplicates.

How to diagnose: Treatment not applied randomly to the units of replication.

Solution: Apply the treatment as feasible, but replicate treatments (incubators) and analyze data based on the appropriate replicates (incubators).

Also, remember that pseudoreplication makes your statistical test more liberal. So if you have

pseudoreplicated and you still did not find a significant effect then don't tie yourself in knots trying to solve the pseudoreplication problem.

Multiple observations subject to the same environmental conditions. For example, you might be interested in the effects of local habitat or relatedness on juvenile dispersal. One potential problem is that all juveniles from the same nest will experience the same local environment. So using multiple individuals from the same nest would be a form of pseudoreplication.

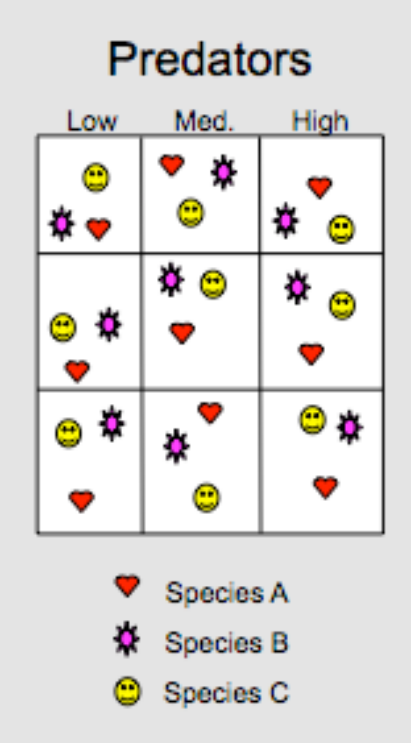
How to diagnose: groups of observations (in this case nest-mates) have the same values for covariates. In an x-y plot with continuous predictors and response variables the data points will appear stacked on top of each other in vertical columns (i.e. for the same covariate value).

Data from the same individual show up in several places in the dataset. For example, a hypothesis about the effects of food abundance on clutch size might be tested using observations across multiple years. In this case some females might be measured in multiple years.

How to diagnose: error df in analysis is > the number of individuals studied when inferences are to be drawn about individuals.

Multiple responses from the same treatment used as independent replicates. This is a bit of a hard one to explain other than through an example. Let's say you were interested in the effects of predator density on the abundance of three species of prey. Ideally you would use a multivariate approach that reflects the fact that you really have three response variables (Species A, B and C). You could potentially do three separate ANOVAs (one for each species), but there is a problem of multiple tests here (we will talk about this later).

What you certainly CANNOT do is to treat the response of each species as a separate replicate (i.e. n = 27) and test a two-way ANOVA using species and predator density as the two factors! This is because the response of one species in a particular replicate



environment (fish tank, enclosure, etc.) is certainly not independent of the abundance of the other species.

Phylogenetic independence in comparative studies. When making interspecific comparisons, individual species' values are not independent replicates since species have shared evolutionary history (Felsenstein, 1985, Harvey & Pagel, 1991). More closely related species are more likely to be similar in body size, life history etc. than more distantly related species. You must correct for this lack of independence.

How to diagnose: Comparative study that uses species as replicates without making use of some phylogenetic method to account for shared evolutionary history.

As an aside the exact same criticism could be applied to the fact that related individuals are not independent of one another. So while it is widely acknowledged that interspecific comparisons need to account for shared evolutionary history, there is very little attention paid to the lack of independence in wild populations due to relatedness.

Where things get really complicated

It is all fine and good to say that replicates need to be assigned and manipulated at the level at which inferences are to be made and that manipulations ought to be replicated at the appropriate biological level, but what happens when we want to test multiple hypotheses across multiple scales? For example, what if we wanted to know the effects of:

Forest type (mens.; hardwood vs coniferous)

Precipitation (mens.)

Leaf species (man.)

Microbial community richness (mens.)

Bag material (man.)

… on decomposition rates in leaf litter bags placed in the soil for 5 years. In this case the various hypothesized effects are occurring at different scales and would be best tested at different scales. This makes a single best design very difficult!

Drawing Conclusions

Pseudoreplication is often a problem with the scale at which conclusions are drawn. Remember that conclusions are limited to the population from which samples were drawn and that our replicates are only ever independent within the scope of our project (i.e. sampling frame). This must be clear.

Pseudoreplication Summary

- You need to understand it to avoid it where possible
- When it is inevitable, you must acknowledge it and be aware of its consequences
- Pseudoreplication is only an issue with respect to the conclusions that you draw