CHAPTER OVERVIEW

5: Experimental Design

Introduction

During this course, you will learn about statistics, yes, but my hope and goal for your experience in this class is much more than that: **statistical reasoning**. To have statistical reasoning skills, you need to be comfortable with the context of how data are acquired, i.e., **data acquisition**. It's not just about knowing how an instrument performs, e.g., under or over the **limits of quantification** characteristics of the instrument lead to censored **missing values**, missing not at random (MNAR). In a broad stoke view, data are obtained from two kinds of studies: **observational studies** and **experimental** (manipulative) **studies**. what a correctly designed experiment can tell you about the world, and how a poorly designed experiment works against you. At the end of the semester, you should be familiar with the issues of Randomization, Control, Independence, and Replication.

The basics explained

Experimental design is a discipline within statistics concerned with the analysis and design of experiments. Design is intended to help research create experiments such that cause and effect can be established from tests of the hypothesis. We introduced elements of experimental design in Chapter 2.4. Here, we expand our discussion of experimental design.

We begin our discussion with review of definitions and an outline of an important concept in statistics. First, we need to be clear about why we measure or conduct experiments. We do so to collect data (datum is the singular) about a characteristic or trait from a population. Data are observations: they include observations and measurements from instruments. At their best, data are the "facts" of science.

Measurement, Observation, Variables, Values

Measurement is how we as scientists acquire our data (Chapter 3.4). The process of measuring involves the assignment of numbers, codes, or labels to observations according to rules established prior to any data collection (Stevens 1946, Houle et al 2011). **Observations** refer to the units of measurement, whereas **variables** are the characteristics or traits that are measured. A **value** refers to the particular number, score, or label assigned to a particular sample for a variable. Variables are generic, the thing being measured, values are specific to the subject or sample being measured for the variable. Each discipline in biology has its own set of variables and samples may or may not have different values for each variable measured. Variables are summarized as a statistic (e.g., the sample mean), which is a number taken to estimate a **parameter**, which pertains to the population. Variables and parameters in statistics were discussed in Chapter 3.4. Because numbers or scores or labels can be assigned according to different rules, this means that variables may be measured on different kinds of scales or data types. The different kinds of data types were presented in Chapter 3.1.

Missing values

A missing value refers to lack of a value for an observation or variable. Missing values can affect analysis and many R algorithms are sensitive or may fail to run in the presence of missing values. **Censored values** include observations for which only partial information is available. Missing data may be of three kinds, and one of them, **missing not at random** (MNAR), can influence the analysis. MNAR implies some observations are missing because of a **systematic bias**. Instrument **limits of quantification** are an example of systematic bias — for example, spectrophotometric absorbance readings of zero for a colorimetric assay (e.g., Bradford protein assay) may not represent complete absence of the target, but rather, the lower detection limits of the instrument or the assay method (0.1 mg protein/mL in this case).

The other kinds of missing values are **missing completely at random** (MCAR) and **missing at random** (MAR). MCAR implies there is no association between any element of the experiment and the absence of a value. Analysis on MCAR data sets may support unbiased conclusions. MAR includes the random errors that occur during data acquisition: date may be lost because of operator error. Analysis of MAR data sets, as with MCAR data sets, may still result in unbiased conclusions; obviously, the size of the data set influences whether this claim holds. In some cases, missing values can be replaced by **imputed values**.

Cause & Effect

Observations or measurements gathered under controlled conditions (experiments) are essential if we are to answer questions about populations, to separate **cause and effect**, where one or more events is directly the result of another factor, as opposed to **anecdote**, a story about an event which, by itself, cannot be used to distinguish cause from association (e.g., spurious correlations, see Ch 16.2). Recall that anecdotal evidence typically comes from personal experience, where observations may be obtained by non-systematic methods. Well-designed experiments, in the classical sense, permit discrimination among competing hypotheses in large part because observations are collected according to strict rules. Well-developed hypotheses tested by well-designed experiments permit ruling out alternative explanations (sensu Pratt [1964] **strong inference**). Observational studies, or epidemiology studies if we are talking about investigations of risk assessment, also may contribute to discussions of cause and effect (see history of smoking by Doll 1998).

Medicine is replete with stories about how a patient showed a particular set of symptoms, and how a physician applied a set of diagnostic protocols. An **outcome** was achieved, and the physician reports the outcome and circumstances related to the patient to her colleagues. This is an example of a **case study**, and the focus of investigation is the individual, the patient. The doctor's report will sound like, "I tried the standard treatment given the diagnosis, and the patient's symptoms diminished, but later returned. I tried a higher dose, but the patient's symptoms persisted unabated." No inferences are made to a wider set of individuals and the report is anecdotal. If observations are made on several patients, this may be a **case series**.

In ecology, a field biologist may notice a six-legged adult frog (Scott 1999, Alvarez et al 2021) — since frogs typically have four legs, the six-legged frog attracts the biologist's eye and he jots down the circumstances in which the frog were found: relative humidity, air temperature, ground temperature, where the frog was found (on lower leaf of a philodendron plant). A water source near where the frog was found is tested for pH with a meter the biologist carries, and a water sample is taken for later testing of herbicides. The frog is collected so that it can be checked for skin parasites. Upon further inspection, the frog did indeed have parasites known to cause deformities in other frog species. However, note that this example too, is a case study. Although the biologist makes additional observations, any conclusions about why the frog has six legs is anecdotal.

From these case studies, no conclusions can be drawn. We cannot say why the patient failed to respond to treatment, nor can we say why the frog has six legs. Why? Because these are singular events, and a variety of explanations can be given as to their causes — importantly, no controls are available, so there's no way to distinguish among possibilities.

From such anecdotes, however, experiments can be designed. The physician may decide to recruit additional patients with the diagnosed illness and apply the standard treatment to see if her anecdote is a single, unique event, or more indicative of a problem with the treatment. The biologist may collect other frogs from the area near where he found the six-legged frog and check to see if they, too, have the parasites. If additional patients fail to respond to the treatment, then the singular even is more likely to be a phenomenon. If the normal frogs also have similar levels of the parasite then it is unlikely that these parasites caused the malformed frog. With this simple step (recruiting similar patients, finding additional frogs), we can begin to make inferences about cause and effect and in some cases, to generalize our findings.

This is the objective of most statistical procedures, the concept of sampling from a reference population and making distinctions between groups within the sample. The difference between observational and experimental studies then is how the subjects are selected with respect to the groups. In an experiment, the researcher controls and decides which subject receives the treatment; therefore, allocation to groups is manipulated by the researcher. In contrast, subjects included in an observational study have already been "assigned" to a group, but not by us. Assignment to groups such as smokers or non smokers, Type II diabetes or no diabetes, etc. is done by nature.

Now, I do not wish to imply that research that cannot be generalized back to a reference population are worthless. Far from it. In fact, there is a strong argument for specificity. For example, much basic biological research depends on work in model organisms, which in turn may be further partitioned into specific genetic lines (cf. discussion in Rothman et al 2013). And my goodness, what we have learned about the devastation to oceanic islands like Guam when the brown tree snake was introduced (Fritts and Rodda 1998). Strictly speaking, what has happened to Guam is a case history. But no one would argue that what has happened to Guam cannot happen to Hawaii and other oceanic islands (e.g., United States Federal law 384-108 "Brown Tree Snake Control and Eradication Act of 2004"). In other words, even from case histories, generalizations can sometimes be made.

There can also be real reasons to ignore the issue of generality. One benefit of specificity is experimental control. Transgenetic lines may differ by single gene knockout or by gene duplication, and clearly the aim of such studies is to evaluate the function (hence purpose) of the gene product (or its absence) on some phenotype. In this sense it may not seem important that a transgenic

mouse is not representative of a wild outbred mouse population. However, this argument is fundamentally one of expedience — such studies do result in specific results, results that cannot be generalized beyond the strains involved. It ignores the issue of genetic background — all of the genes that affect a trait in addition to the candidate gene under study (Sigmund 2000; Lariviere et al 2001). Transgenic mice of different inbred strains or their hybrids may have very different alleles at other genes that may influence a phenotype; hence, the results of the gene knockdown or other engineering result in different outcomes. Results of genetic manipulations on inbred strains, no matter how sophisticated, mean that the conclusions are strain- or hybrid cross-specific. Thus, although technically and financially difficult, conclusions are better, more **generalizable**, when conducted with many different inbred lines and verified in outbred mouse populations precisely because genetic background often influences function of single genes (Sigmund 2000; Lariviere et al 2001).

"Causal criteria:" Logic of causation in medicine

This section is in progress. Just a list of key points and references

Throughout this text, emphasis on the power of experimentation is emphasized. Well-designed experiments should consider …

Henle-Koch's postulates (1877, 1882), developed from working on tuberculosis to report a set of causal criteria to establish link between a microorganism and a disease, are the following:

- The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.
- The microorganism must be isolated from a diseased organism and grown in pure culture.
- The cultured microorganism should cause disease when introduced into a healthy organism.
- The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

Robert Koch wrote these more than 100 years ago, so, clearly, understanding of infectious disease has improved. **Evan's postulates**, quoted from *A Dictionary of Epidemiology*, 5th edition (pp. 86-87):

1. Prevalence of the disease should be significantly higher in those exposed to the hypothesized cause than in controls not so exposed.
2. Exposure to the hypothesized cause should be more frequent among those with the disease than in controls without the disease —when all other risk factors are held constant.
3. Incidence of the disease should be significantly higher in those exposed to the hypothesized cause than in those not so exposed, as shown by prospective studies.
4. The disease should follow exposure to the hypothesized causative agent with a normal or log-normal distribution of incubation periods.
5. A spectrum of host responses should follow exposure to the hypothesized agent along a logical biological gradient from mild to severe.
6. A measurable host response following exposure to the hypothesized cause should have a high probability of appearing in those lacking this before exposure (e.g., antibody, cancer cells) or should increase in magnitude if present before exposure. This response pattern should occur infrequently in persons not so exposed.
7. Experimental reproduction of the disease should occur more frequently in animals or humans appropriately exposed to the hypothesized cause than in those not so exposed; this exposure may be deliberate in volunteers, experimentally induced in the laboratory, or may represent a regulation of natural exposure.
8. Elimination or modification of the hypothesized cause should decrease the incidence of the disease (e.g., attenuation of a virus, removal of tar from cigarettes).
9. Prevention or modification of the host's response on exposure to the hypothesized cause should decrease or eliminate the disease (e.g., immunization, drugs to lower cholesterol, specific lymphocyte transfer factor in cancer).
10. All of the relationships and findings should make biological and epidemiological sense.

Fredericks, D. N., & Relman, D. A. (1996). Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clinical microbiology reviews*, *9*(1), 18-33.

Correlation (association) does not imply causation, a well-worn truism in any application of critical thinking skills.

Note:

Association is the more general term for a possible relationship between two or more variables. A correlation in statistics generally refers to a linear association (Chapter 16); the aforementioned truism should be restated as association does not imply causation.

However, sometimes association does point to a cause. A familiar example is association between tobacco cigarette smoking causes lung cancer. Surgeon General Luthar Terry's 1964 report (link to document in National Library of Medicine) presented a strong case linking smoking to elevated risk of lung cancer and coronary artery disease.

Bradford Hill's guidelines to evaluate causal effects based on epidemiology (Hill 1965, see also Sussar 1999, Fedak et al 2015). They form a set of necessary and sufficient conditions, based on inductive reasoning.

1. Strength of association
2. Consistency of observed association
3. Specificity of association
4. Temporal relationship of the association
5. Biological gradient, e.g., a dose-response curve
6. Biological plausibility
7. Coherence, the cause and effect inference should not conflict with what is known about the etiology of a disease.

Follows and extends David Hume's (1739) causation criteria: association (Hill #1), cause precedes effect (Hill #4), direction of connection.

Validity

An obvious objective of research is to reach valid conclusions about fundamental questions. A helpful distinction between the specific and the generalizable experiment is to recognize there are two forms of validity in research: **internal validity** and **external validity** (Elwood 2013). Internal validity is the quality of a designed study that determines whether cause and effect can be determined. Random assignment of subjects to treatment groups enhances the internal validity of the study. External validity relates to how general the assessment of cause and effect can be to other populations. Thus, random sampling from a reference population has to do with whether or not the study has external validity.

Additional definitions

We proceed now with definitions. We use the term **population** in a special and restrictive way in statistics. Our definition includes the one you are already familiar with, but it also means more than that.

Populations are the entire group of individuals that you want to investigate. In statistics, the entire groups is actually the entire class with the observation — so if we are referring to the average body weight of house mice, we're actually referring to the body weight as the population — it's a subtle distinction, not essential for our introduction to biostatistics. When we conduct experiments and apply statistical tests on collected data, we generally intend to make inferences (draw conclusions) from our results back to the population.

Population has a strict application in statistics, but the definition also includes our general understanding of the word population. For example, examples of a population in the general sense that one may refer to include:

- the entire human population existing today.
- the entire collection of U.S. citizens.
- all the individuals in an entire species.
- all individuals in a population of a species (e.g., house mice in a dairy barn in Hawai'i).
- all of us in this class room (if we are *only* interested in *us*).

If you could measure the entire population then there would be no need to do (or learn) statistics! Populations usually are in the thousands, millions, or billions of individuals. Here, population is used in the everyday sense that we think of — a collection of individuals that share a characteristic.

A more formal definition of "population" in statistics reads as follows: A statistical population is the complete set of possible measurements on a trait or characteristic corresponding to the entire collection of sampling units for which conclusions are intended.

To conclude, in this class, when we talk about population, we will generally be using it in the everyday sense of the word. However, keep in mind that the definition is more restrictive than that and the key is to identify what sampling units are measured.

Conclusions

This is only the beginning, the basics of experimental design. Entire books are written on the subject, as you can well imagine. We will also return to the subject of Experimental Design throughout the book. We will return to random sampling in Chapter 5.5. Next we discuss distinctions between experiments and observational studies with respect to sampling of populations.

A bit of a disclaimer here before proceeding; while I cite several papers for examples in experimental design in Chapter 5, readers should not read into this that I am either criticizing or endorsing the published experiments. Experimental design will always have elements of compromise — the trick, of course, is knowing which choices influence validity (Thompson and Panacek 2006).

Questions

1. Define in your own words the following terms:
   - reference population
   - subjects
   - specific versus general conclusions
   - random sampling
   - convenience sampling
   - haphazard sampling
   - research validity
2. Revisit our cell experiment, “What is the sampling unit in the following cell experiment?” How would you change this experiment so that there will be biological and not just technical replication?
3. Describe the type of sampling for each research scenario described:
   - All African snails on a staircase at Chaminade University are collected on a Thursday evening.
   - All African snails on a staircase at Chaminade University are collected every Thursday evening for six months.
   - All African snails on all staircases at Chaminade University are collected.
   - African snails are studied in the lab, then returned to the areas from which they were collected. Days later, the researcher collects snails from the same area.
   - African snails are studied in the lab, then returned to the areas from which they were collected. Days later, the researcher collects snails from a different area.
   - The Chaminade University campus is divided into grids. Grids that include stairwells are marked. Before collecting snails, the researcher randomly selects from the list of grids and searches for snails only in those grids selected from the list.

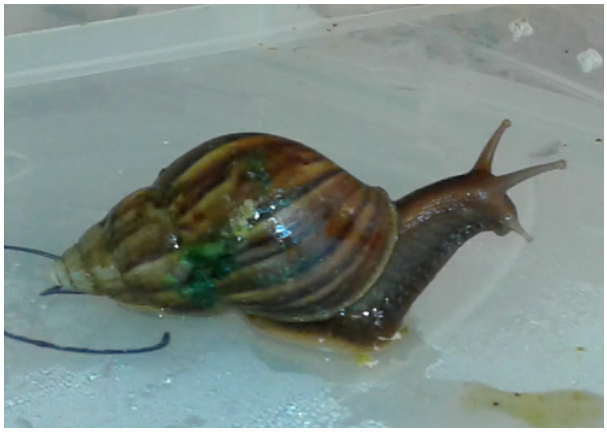

Figure 5.1: Giant African Snail (*Lissachatina fulica*, formerly *Achatina fulica*). Image by M. Dohm.

4. A researcher wishes to study the effects of salt on mosquito larval survival. He works with *Aedes* species, mosquitos that are characterized as “container-breeding” – their larvae develop where water accumulates in tree holes or indentations in rock, or even in the containers left by humans (e.g., tires, flower vases or planters). His preliminary experiment is outlined in the following table. The last column indicates the measurement that he plans to record. Identify the sampling unit. Identify the experimental unit

5. Consider the Hermon Bumpus House sparrow survival data set (described at Field Museum (Chicago, IL) and American Ornithology Society), famous as an early example of natural selection. A storm on 1 February 1898, in Providence Rhode Island left dozens of house sparrows on the ground. Birds were collected and brought to Bumpus's, 136 in all. Seventy two revived, 64 died. Bumpus identified the sex and measured nine morphological traits of each bird. Bumpus concluded from his graphs that males survived better than females and that shorter, lighter birds with longer legs, wings and sternums and larger brain size ("skull width") also survived better. Which type of study is the Bumpus study? Select one:
   - Case study
   - Anecdote study
   - Case control study
   - Cohort study
   - Cross-sectional study
6. This next scenario may be evaluated by you for potential sources of bias. Review the list of bias listed above. A researcher wants to do a population count of feral cats on campus. Feral cats are active at night, so he decides to set up a feeding station near a light post. The researcher sits all night in a parked car yards and watches the feeding station for visits by cats. The researcher repeats these observations over the course of a week, moving the feeding station to different campus locations each night, and reports the total number of cats seen during the week as an estimate of the population size. Be able to discuss this study in terms of potential and actual bias.

5.1: Experiments
5.2: Experimental units and sampling units
5.3: Replication, bias, and nuisance
5.4: Clinical trials
5.5: Importance of randomization
5.6: Sampling from populations
5.7: Chapter 5 References

This page titled 5: Experimental Design is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael R Dohm via source content that was edited to the style and standards of the LibreTexts platform.