Chapter 2

Data

Quantitative methods and statistics are applied to *data*. Data are observations, not things. Data are not artifacts. They are not pots or stones or bones or any other component of the phenomenological world. We build data by making systematic observations on pots and stones and bones. What constitutes data is determined by our research questions and theoretical perspective. In other words, we create data to serve a purpose defined by a pre-existing intellectual framework to achieve an end. Most certainly, the real world exists in terms of various arrangements of matter and energy, but that real world is not to be confused with the observations that we make about it.

In addition to the theoretical perspective that we bring to bear and the research question we are addressing, the tools with which we can look at the world also influence what our data look like. As Gulliver's travels taught us, the world looks very different to Lilliputians and Brobdingnags, and the view is very different when the instrument we hold to our eye switches from a telescope to a microscope.

There is no "high court" of archaeologists that make the rules about what kind of observations we are restricted to make, or at what scale. Data are what we determine them to be. Certainly, there is a range of observations that many archaeologists agree to be useful towards addressing particular problems, but by no means are we restricted to looking at the world from a single perspective. When we build data we need to be sure that we do so in a systematic manner. All measurements and observations must have unambiguous definitions and they must be consistently recorded in the same way in order to eliminate measurement bias.

We build data by making observations on *variables*. A variable is any quality of the real world that can be characterized as having alternative states. The color of soil is a variable, as is projectile point length. With the former, the color spectrum is partitioned arbitrarily into segments with labels. The labels can be commonsensical, e.g., reddish brown, or we can use a standardized labeling system that refers to a definition of the label. For color, most archaeologists use Munsell color charts, with standardized labels such as 10YR 4/1. With respect to projectile point length, most archaeologists use the metric system where length is arbitrarily partitioned into usefully sized segments. How the measurement used to characterize a variable is partitioned depends on our theoretical perspective and research problem. For example, our research problem might suggest that millimeters is the perfect sized segment with which to measure the height of a ceramic vessel, but wholly inappropriate for the measurement of the size of a settlement or of one grain of corn pollen. Furthermore, a different analytic framework might not consider ceramic vessels or corn pollen worthy of measurement for addressing the same research problem.

An individual observation of a variable is called a *variate*. If we measure the rim angles of ten ceramic vessels, we have manufactured ten variates. These variates constitute data. Again, note that data are not variables. They are the totality of our observations, or variates. Also note that the word “data” is plural. A single observation is a variate and multiple variates are data. It is consequently improper to refer to data as a singular object; your data are, they are not is.

Scales of Measurement

When we create data our observations can be recorded in one of four measurement scales. The statistical tools available for analysis differ depending upon what measurement scale is being used. These measurement scales are: *nominal*, *ordinal*, *interval*, and *ratio*.

Nominal Level of Measurement

Nominal level of measurement exists either when we are using numbers as labels (hence the term "nominal"), or when we are using numbers to represent the abundance of a class of phenomena (i.e., counts). This can be confusing, but the literature refers to both uses of numbers as being nominal level measurement, so it is best to understand the distinction.

Using numbers as labels is common in the world; one only needs to turn the TV channel to ESPN to see numbers as labels on volleyball players, football players, field hockey players, horses, race cars and innumerable other rapidly moving objects. These numbers constitute labels only, and anyone performing arithmetic operations on football player's jersey numbers should seek professional help immediately.

Archaeologists use numbers or letters quite frequently as labels to code data. In their research, a ceramic analyst might use the label "1" to designate an observation taken on a cross section of a sherd that was oxidized externally but not on the interior. The label

"2" might be used to designate the observation of no oxidation at all. Let us say that a total of eight numbers are used to designate, or label, various attribute states of a qualitative variable. Our fictitious analyst could have used labels such as Sally and Harry almost as easily. These labels have no mathematical significance, and to perform meaningful arithmetic operations on them would be impossible.

There is a distinction, however, between how our analyst was using numbers as labels and the use of numbers to label racecars. While there is no analytic utility to considering the abundance and distribution of race cars labeled "1" through "8" across racetracks, a researcher might be very much interested in the abundance and distribution of eight classes of ceramic firing attributes across the landscape. This information can be used to evaluate ideas about differences in technology and site function. Additionally, if variation is present across sites or across site components in terms of class abundance, most analysts will ultimately seek to explain those differences.

A number of good statistical tests exist to determine if differences are meaningful, including the chi-square test, which is one of the first statistical tests used in archaeology (Spaulding 1953) and is still commonly used. In archaeology, nominal level data are attributes of qualitative variables, often attributes of artifacts. These kinds of variables are called *discrete* variables (they are also referred to as *discontinuous* variables, or *meristic* variables). Common discrete variables for different kinds of artifacts include:

- Stone tools: raw material type, the presence or absence of edge wear, type of edge wear, number of flake scars, presence or absence of heat treatment.

- Ceramics: temper type, extent of paste oxidation, type of decoration, type of paint, presence or absence of design elements.
- Ground stone: type of abrasion, direction of abrasion, artifact shape.
- Bone: presence or absence of burning, type of cut marking.

Counting is the only appropriate arithmetic operation to be used on discrete or nominal variables, and the values assigned are always whole numbers. For example, one cannot have 5.5 pieces of Edward's Plateau chert on a site, nor can one have 20.3 Bison bones.

Ordinal Level of Measurement

Ordinal levels of measurement imply an observable order or degree along some defined dimension of a variable. Ordinal measurements are only possible when measuring continuous variables. Continuous variables can logically assume any value and are not limited to simple discrete intervals. For example, length, size, and intelligence are continuous variables. Ordinal measures divide continuous variables into segments that are assumed to be asymmetrical. That is, we cannot assume equal values for our units of measure. Ordinal measures come with labels like small, medium, large, and extra large, where we cannot assume the distance between large and extra large is the same as the distance between small and medium. We either do not have the necessary instrument with which to measure the variable of distance in equal increments, or the phenomena we are measuring cannot be measured in such a manner. This happens often in archaeology, particularly when we are using one kind of data to infer past behavior, or another set of data.

For example, Grayson (e.g., 1984:96-111) has argued that counts of faunal specimens recovered from archeological sites are ordinal counts, at best. His arguments rest on the fact that we often have only bones when we are actually interested in questions related to the relative contributions of different species of animals to the subsistence practices of the people that occupied a settlement. Only in rare circumstances would archaeologists recover all of the bones of all of the animals or even a single bone from each animal utilized at a settlement. As a result, the analysis of the faunal material can at best only indicate which animals did and did not constituted a large portion of the diet, but cannot indicate the absolute number of any animal species consumed. Grayson therefore concludes that both procedures in common use, the Minimum Number of Individuals (MNI) and the Number of Identified Specimens (NISP), produce ordinal measures of abundance. The same holds true for both palynological and ethnobotanical measures as well.

Settlement hierarchies are an additional example of ordinal measurement in archaeology. For example, Di Peso et al. (1974) argued that sites in northern Mexico could be divided into three tiers reflecting their relative size. At the top of the settlement hierarchy was the large site and presumed regional center, Paquimé, which was characterized by multistory construction, public and ceremonial architecture, platform and effigy mounds, and craft specialization. Di Peso and his colleagues estimated more than 2,000 people inhabited Paquimé. Subordinate to Paquimé were sites that were roughly 1/2 to 1/3 the size of Paquimé and that Di Peso believed were occupied by 500 to

1,000 people. Di Peso further argued that smaller sites, which he believed were occupied by 100 to 500 people, were subordinate to both Paquimé and the larger sites. The difference between the first, second, and third tier is not uniform, and each tier does not reflect a consistent unit of size.

In a similar manner, many researchers use rank orders to measure social complexity. Sahlins and Service's typology of bands, tribes, chiefdoms and states is one such example. The distance between bands and tribes is not the same as the distance between chiefdoms and states.

Interval Level of Measurement

With interval level measurements our continuum is partitioned symmetrically into even increments. As we wrote the first draft of this paragraph we were in the middle of conducting fieldwork on the Zuni Reservation in western New Mexico. When the senior author crawled out of his trailer to make coffee this morning he noted that it felt much colder than the previous morning. This was an ordinal level observation, as having no thermometer, he sensed that it was colder, but he had no idea of how much colder it was. In contrast, a thermometer provides an interval level of measurement that is much more useful than his impression of relative coolness. The increments are symmetrical, and the difference between 40 degrees and 50 degrees is the same as the difference between 60 degrees and 70 degrees. The degree of difference between measurements can thus be specified in a way that is not possible when using measurements of an ordinal scale such as warmer and colder.

A common interval level measurement used in archaeology is direction. Using a compass, the directions from a point on the landscape are divided into 360 equal units. Each degree represents a uniform and consistent unit of measurement.

Ratio Level Measurement

Ratio level measurement is the same as interval level measurement, with one exception, the addition of a true zero value. The value 'zero' is not arbitrarily assigned. In the temperature example presented above, zero is an arbitrarily chosen value. Zero degrees does not mean that there is no temperature, as is illustrated by the differences between zero degrees in the Fahrenheit and Celsius scales. Likewise, zero degrees on a compass does not mean an absence of direction. The zero point on the compass is arbitrary and could just as easily be due south or northwest. In contrast, zero length of a piece of debitage means just that, no length at all. Obviously, any piece of debitage we measure must have a length value no matter how minute. Zero length still represents the complete absence of length, though. What is important, then, is that it is theoretically possible to get a value of zero for ratio level measurements, which opens up your analysis to a much broader assortment of mathematical and statistical procedures.

Many statistical applications assume ratio level measurements. Unfortunately, many archaeologists ignore the fact that much of their data is constructed in terms of ordinal level measurements, and perform inappropriate analyses regardless. Such analyses are unfortunately meaningless, which relates to our next topic: *validity*.

Validity

Validity in measurement is ascertained by determining if we are measuring what we think we are. One dictionary provides this definition for valid: "well grounded or justifiable; being at once relevant and meaningful." This seems fairly straightforward: we are either measuring the volume of a pot or we are not. Determining the validity of the measurement in this case is straightforward. Other situations are no so clear-cut, however. Archaeologists are seldom interested in measures of material items or pieces of technology, per se. Most are actually interested in what those measures on technology tell us about past human behavior. For example, it might come to mind to measure pot volume to look at the importance of storage behavior as it relates to agricultural intensification in two different time periods in the North American Southwest. One must then ask the question, “Is pot volume a measure of the importance of storage?” Well, one pot no. But perhaps if we total all of the volume measurements of all of the whole pots recovered from each of the time periods, and compare them, perhaps that will tell us during which period people were more dependent upon storage. But then again, maybe not. What about all of the sherds? Can we take them into account? Certainly there are many more sherds than whole pots, but sherds are broken, making volume measures difficult. Perhaps we can estimate volume with broken sherds, or assume that the whole pot/sherd ratio is the same across time periods. But one time period is older, and sherds had a longer time to break into smaller pieces. And what if we have different numbers of sites in the different time periods? And what if the time periods are of different lengths? What are the implications of all of these considerations?

The main implication is that determining validity is not quite as simple as it seems. Can pot volume be used as a measure of the importance of storage? That is, is it "well grounded or justifiable; being at once relevant and meaningful"? Probably, but a logical argument has to be made for the validity of the measurement. Importantly, the establishment of validity is a logical argument, not a mathematical one, or one made by reference to authority.

Another good example to consider with respect to validity is the measurement of population size. Clearly, population numbers are important to determine for many questions archaeologists ask, yet, our populations are long gone. What surrogate measures can we use? A wide variety has been offered, but questions can be raised about the validity of each. In the areas in which we have worked, room numbers, site numbers, habitation areas, and site areas, among other measures, have been used to estimate population numbers. Many arguments have been made for and against the validity of each, which is very healthy for our discipline. The continued assessment of validity is one of the functions of the scientific enterprise.

In many instances, especially where we are restricted to surrogate measures of what interests us, we have to draw the conclusion that our measures are correlated sufficiently with our target data set that we are willing to proceed. A good example comes from dendroclimatology. There is a well-established correlation in many tree species between tree-ring width and the amount of water available. Therefore, given that we don't have a time machine to travel to the time period that interests us to set out rain gauges, we use the widths of the rings of trees recovered in archaeological sites as a surrogate measure for what we want to measure, our target data set, moisture. Despite a number of criticisms, enough arguments have been presented regarding the validity of the measure that most archaeologists are quite comfortable with it.

The battle isn't over, though, just because validity has been ascertained. We must focus on the *accuracy* and *precision* of our instruments. We use the concepts accuracy and precision in the way proposed by Sokal and Rohlf (1981:13), "Accuracy is the closeness of a measured or computed value to its true value; precision is the closeness of repeated measurements of the same quantity." The terms thus are not synonymous. A biased instrument could produce precise but inaccurate values. In general, we wish to attain both accuracy and precision. An Electronic Distance Measurer (EDM) is both more accurate and precise than an optical transit, but we are still quite comfortable with maps generated by optical transits, suggesting that error isn't as important here as convenience. The EDM simply shoots farther than the optical, and automatically calculates true distance without putting the users through the horror of remembering high-school trigonometry. This implies that the EDM supplies accuracy and precision greater than we actually need, which is an enviable position unless the greater accuracy and precision costs more than it is worth.

Populations and Samples

When we make observations on the world we can make observations at the scale of a population or of a sample of a population. A *population* is the totality of phenomena about which we wish to draw conclusions. A *sample* is some subset of that population. Populations and samples are finite in time and space, and are defined by the investigator. Ideally, we would like to measure the totality of the phenomena about which we wish to draw inferences. However, that is often expensive, unnecessary, or impossible.

A researcher may wish to draw conclusions about the physical characteristics of Achulean handaxes without making observations on the totality ever excavated. We may consider the total excavated as the population, and look at only a small subset of them to draw conclusions about the population as a whole. We use statistics to use information gathered on samples to make inferences about the population. Note that we defined the population as all Achulean handaxes ever excavated. Our imaginary researcher, however, is probably interested in more than those excavated. He or she is probably interested in all Achulean handaxes ever used, and this is probably really the more accurate population of interest. All of the Achulean handaxes ever excavated is actually a sample of all of the Achulean handaxes ever used because it is unlikely that we have recovered every single Achulean handaxe. Therefore, we sample. Thankfully, statistical techniques are available that allow us to make useful inferences about the population from our sample. In fact, drawing inferences that characterize a population from a sample is one of the main purposes of statistics.

What constitutes an adequate sample size is case by case specific, and procedures will be introduced later on in the book that will allow you to make such determinations. In general, however, the more homogeneous the population is, the smaller your sample size can be. A physician can draw inferences concerning your blood chemistry from a single blood sample because blood is a homogeneous fluid that does not differ from body part to body part. Much of the archaeological record is similarly redundant, while other segments are extremely heterogeneous.

So, how does one work from a sample to a population? This is the subject matter of the remainder of the text.