FREQUENCY DISTRIBUTIONS

A. What is a frequency distribution?

The first kind of univariate, descriptive statistical technique that we will use is the frequency distribution. This table is a commonly used graphic display; we'll talk about several others later in the course.

Katzer et al. (1998, p. 272): a frequency distribution is "A table that shows how frequently each value of a variable occurs in a set of scores."

What is a frequency distribution (cont'd)

An ordered array of data points (also called observations, measures, measurements, scores) from highest to lowest or vice versa

For the purposes of this class, we will always generate frequency distributions from highest to lowest value.

Before we can look at frequency distributions, we need to discuss two important concepts in research methods: variables and cases.

B. What is a variable?

Katzer et al. (1998, p. 280): "An attribute or property of a person, group, or object that can be measured; the resulting measures are capable of assuming different values."

Vogt (1999, p. 302): "any finding [or attribute or characteristic] that can change, that can vary, or that can be expressed as more tan one value or in various values or categories. The opposite of a variable is a constant."

Lossee & Worley (1993, p. 106): "An entity that can hold two or more values."

4. What is a case?

While this is a very complex question, for our purposes, a case is an individual person, book, database, library, time period, and so on that you can observe.

Cases are sometimes called units of analysis; we'll discuss that more later in the semester.

Thus, a variable is a characteristic of a case that you can measure, i.e., observe systematically.

5. There are, therefore, three important elements of a variable

An attribute or characteristic of a case (this usage is common, but differs from that in Babbie's text)

The attribute can be measured, that is, it can be observed systematically, **not necessarily quantified**, e.g., eye color and gender

It can vary.

6. Examples abound, but variables are usually classified as either qualitative or quantitative.

The difference between these two types of variables results from how variables are measured **NOT** because of some supposed intrinsic characteristic any variable. Remember that measurement means observation in some systematic way

For example, we can measure (i.e., observe systematically) height by inches or by classifying people immediately into small, medium,. or large. Thus, height is defined by the researcher as, respectively, a quantitative or qualitative variable.

C. An example of a frequency distribution

1. We'll start with some raw data, i.e., direct observations, of values that a variable assumed in a research study

2. A variable is often referred to as "x," a printed, lower case, Roman letter, **not** as "χ," which is a lower case, Greek, script letter – be sure in your notes and exams to write "x" when you mean "x."

3. We will look at a population, as defined by the researcher (we will discuss the concepts of populations and samples a little later in the semester) consisting of 17 values of the variable.

An example of a frequency distribution (cont'd)

4. **x is defined as the number of monographs on Reserves for the faculty members in a department of classics in one semester.**

5. **The values the variable assumes, in the order in which they were determined are below. Remember that all of these values are measured in units of monographs:**

10	5	5	0
9	0	10	5
7	0	9	
23	7	5	
7	0	1	

6. Let's generate the frequency distribution of the variable in this population, ranging from highest value to lowest

We will use five columns in the frequency distribution, and this model will be used throughout the semester:

x, the number of monographs

freq x, the frequency of x, how often this particular value appears in the data set

cum freq x, the cumulative frequency of x, the number of observations in the data set at this particular value or lower

6. Generate the frequency distribution (cont'd)

rel freq x, the relative frequency of x, the proportion of observations that assume this value; this proportion is determined algebraically, by dividing the frequency of x (freq x) for each value by the number of observations (N). Recall that N is the number of observations in a population, while n is the number of observations in a sample.

cum rel freq, the cumulative relative frequency of x, the proportion of scores at this value or lower; this value is only approximately equal to the cumulative frequency of x (cum freq x) divided by N.