**Frequency Distributions
G&W Ch. 2**

**Parameters and Statistics**

When describing data, it is necessary to distinguish whether the data come from a population or a sample.

Typically, every population parameter has a corresponding sample statistic.

- Parameter—a value that describes a population

- Statistic—a value that describes a sample

**Descriptive Statistics**

- techniques used to summarize, organize, and simplify data

- can't look at it all - get a quick, good impression

**Inferential Statistics**

-techniques used to study samples and then make generalizations about the populations from which they were selected

**Variables**

discrete - separate categories. No values can exist between two neighboring categories (e.g., dice)

continuous - infinite fineness. There are an infinite number of possible values that fall between any two observed values (e.g., time)

- each score corresponds to an interval of the scale

- the boundaries that separate these intervals are called real limits

**Frequency Distribution**

Goal: simplify the organization and presentation of data

Definition: lists or displays graphically the number of individuals located in each category on the scale of measurement

-  takes a disorganized set of scores and places them in order from highest to lowest, grouping together all individuals who have the same score

-  see the set of scores “at a glance”

-  frequency distribution can be structured either as a table or graph but both display (1) all categories that made up the measurement scale and (2) the number of individuals in each category

**Frequency Distribution Tables**

data: 3,1,1,2,5,4,4,5,3,5,3,2,3,4,3,3,4,3,2,3

-  *X*—categories of the measurement scale (usually listed highest to lowest). To calculate the sum of scores, must use both X and f columns

-  *f*—frequency, number of individuals in that category

- To obtain the total number of individuals in the data set, add up the frequencies 

-  *p*—proportion, the proportion of the total number of responses that fall into this category  (*p = f/N*)

Cumulative frequencies *(cf)* show the number of individuals located at or below each score

- Cumulative percentages *(c%)* show the percentage of individuals accumulated as move up the scale

 *x f p % cf c%*

 5 3 .15 15% 20 100%

 4 4 .20 20% 17 85%

 3 8 .40 40% 13 65%

 2 3 .15 15% 5 25%

 1 2 .10 10% 2 10%

Grouped frequency distribution tables—group the scores into intervals and list these intervals in the frequency distribution table. The wider the interval, the more information that is lost. Should have about 5-10 intervals depending on range of data. Width of each interval should be an easy number (5 or 10) and all intervals should be the same width.

rank or percentile rank—the percentage of individuals in the distribution with scores at or below the particular value.

**Frequency distribution graphs/charts**

- x-axis (abscissa)—lists the measurement scale categories

- y-axis (ordinate)—lists the frequencies

histogram—a bar is drawn above each X value, so that the height of the bar corresponds to the frequency of the score. If data is from interval or ratio scale, the bars are draw on so that adjacent bars touch each other. The touching bars produce a continuous figure, which emphasizes the continuity of the variable. (SPSS? yes)

bar graph—like a histogram, a bar is drawn above each X value, so that the height of the bar corresponds to the frequency of the score. If data is from nominal or ordinal scale, graph is constructed with space between the bars. (SPSS? yes)

frequency distribution polygon—used with interval or ratio scales instead of a histogram. A single dot is drawn above each score so that the height of the dot corresponds to the frequency. (SPSS? no)



  



**The Shape of a Frequency Distribution**



**Measures of Shape of a Distribution**

**1.  Skewness**

Measures to what extent a distribution of values deviates from symmetry around the mean. A value of zero represents a symmetrical or balanced distribution. Skewness values between +/- 1.0 are considered excellent for most purposes, but values between +/- 2.0 are also acceptable in many cases, especially for basic research.

Positive skew--Tail is in the positive direction.  There are fewer larger scores than we would expect with a normal distribution.

Negative skew--Tail is in the negative direction.  There are fewer smaller scores than we would expect with a normal distribution.

**2.  Kurtosis**

Measures the "peakedness" or "flatness" of a distribution. A value of zero indicates the shape is close to a normal distribution.  A positive value indicates a distribution is more peaked than normal.  A negative value indicates a distribution is flatter than normal.  As with skewness, a value of zero represents a distribution that is shaped very similarly to a normal distribution. Kurtosis values between +/- 1.0 are considered excellent for most purposes, but values between +/- 2.0 are also acceptable in many cases, especially for basic research.