Descriptive Statistics

Individuals are described in terms of their characteristics. When characteristics are measured for a group of individuals, summaries of the measures are called descriptive statistics. The measure of the characteristic of individuals in a group is called the **dependent** variable. When groups are compared the variable that represents the differences in the groups (i.e., gender) is called the **independent** variable.

**Measurement Scales**

When data are collected from individuals in a group, the scale used for data collection can be of three (four) types.

Nominal data—data sorted into named containers and the order of the containers has no meaning.

Ordinal data—data sorted into containers in which the containers have a natural order but there is no way to know if the containers are evenly spaced.

Interval data—data sorted into containers that have a natural order and the distance between containers is consistent throughout the scale—a ruler.

(Ratio data)—ratio scales are interval in nature but the scale starts at zero. This allows *ratio statements* to be made about the data.

# Frequency Distribution

A frequency distribution counts the number of times that a score occurs at each level of a variable (characteristic). A graph of a frequency distribution is called a *histogram*.

# Measures of Central Tendency

Mode—the score value that occurs most frequently in the distribution (the only choice for nominal data).

Median—the score value that divides the distribution into the lower and upper 50 percent of the scores (best for ordinal data, although a mode could be used).

Mean—the “center of gravity” of the distribution, such that the scores above the mean exactly balance the scores below—the average (best for interval data, although a median or mode could be used).

# Measures of Variability

Range—the highest score minus the lowest score (sometimes the range is reported as the actual highest and lowest values).

Standard Deviation—is determined computing the absolute value of the distance each value is from the mean and then averaging the differences (using degrees of freedom: *df*) computed. The standard deviation is expressed as SD or σ (sigma) or sometimes s.

**Standard Curves (Normal Distribution)**

Symmetrical

Mean = Median = Mode

SD determines width of curve

Percentage relative to SD remains the same

1SD = 34.13 (68.26)

2SD = 47.72 (95.44)

3SD = 49.86 (99.27)

# Standard Scores

*z* Scores—are the distance each score is from the mean measured in number of standard deviations. *z* scores can be translated into percentile rankings.

# Correlations

Correlations show the strength and direction of the relationship of two variables in the sample. Correlations are expressed on a continuum from 1 to -1.

• A correlation between .7 and 1 (or -.7 and -1) are called strong correlations.

• A correlation between .3 and .7 (or -.3 and -.7) is called a moderate correlation.

• A correlation between zero and .3 (or 0 and -.3) is called a weak correlation.

• A minus sign in a correlation coefficient indicates that the two variables behave in the opposite direction.

• Pearson’s *r*—is a test of correlation used when both variables are interval/ratio.

• Spearman’s Rho—is a test of correlation when at least one variable is ordinal.

Inferential Statistics

A **population** is the largest group to which you wish to generalize.

A **parameter** is a summary of a measure of a characteristic of a population.

A **sample** is a randomly selected subgroup of a population (the people from whom you gather data).

A **statistic** is a summary of a measure of a characteristic of a sample.

Inferential statistics are designed to make inferences about population parameters using sample statistics.

Sampling Distribution of the Mean

The distribution of the means of random samples of a specific size from a population.

Standard Error of the Mean

The standard deviation of the sampling distribution of the mean.

Systematic Error

Sampling Bias

Measurement Bias

Random Error (Sampling Error)

The natural variation of sample means from the population mean.

Significance

A difference between two sample means is said to be significant if the probability that the difference would occur randomly is below a predetermined level.

If the mean and standard deviation of a population are known then a *z* test can be used to determine if a sample is statistically different than a population (one sample *t*).

If the mean and standard deviation of a population are not known then statistics from samples of a population can be used as estimates. In this case an independent samples *t* test is used to determine significance.

If two groups are being compared and exactly the same individuals are in both groups then a paired samples *t*-test can be used.

One-Tailed and Two-Tailed Tests

When comparing samples of a population, if the likely direction of a difference can be firmly established then a One-Tailed test of significance can be used.

When comparing samples of a population, if the likely direction of a difference cannot be firmly established then a Two-Tailed test of significance is used.

If multiple group comparisons are being done simultaneously then the appropriate test is an ANOVA.