

Glossary of Key Data Analysis Terms

Levels of data

Nominal Variable - A variable determined by categories which cannot be ordered, e.g., gender and color.

Ordinal Variable - A variable in which the order of data points can be determined but not the distance between data points, e.g., letter grades and extent of agreement.

Interval Variable - A variable in which both order of data points and distance between data points can be determined, but for which no real zero point exists, e.g., temperature.

Ratio Variable - A variable in which both order of data points and distance between data points can be determined, and for which there is a real zero point, such as number of drinks, weight, or distance.

Describing the data set

Central Tendency- These measures indicate the middle or center of a distribution.

Descriptive statistics - Analyses designed to describe and/or summarize the data set.

Distribution - The range of values of a particular variable.

Deviation - The distance between the mean and a particular data point in a given distribution.

Frequency table - A way of summarizing data that shows how many times each value (or set of values) of the given variable occurs. The percentage/proportion of times each value occurs may also be recorded.

Mean - The average score within a distribution.

Median - The center score in a distribution.

Mode - The most frequent score in a distribution.

Normal distribution - A frequency distribution, depicted by the bell curve, that represents the probability that a majority of randomly selected members of a population will fall within the middle of the distribution. Many statistical tests are based on the assumption that the population's data on a given variable would comprise a normal distribution.

Range - The difference between the highest and lowest scores in a distribution.

Standard Deviation - A measure of variation that indicates the typical distance between the scores of a distribution and the mean.

Variance - A measure of variation within a distribution. It is based on the deviations of individual scores from the mean.

Identifying relationships between variables

ANOVA (Analysis of Variance) - A method of statistical analysis used to determine differences among the means of two or more groups on a variable. The independent variables are usually nominal, and the dependent variable is usually interval.

Causal Relationship - An association between two variables in which one causes a change in the other.

Chi-square test- A statistical test commonly used to compare observed data with the data that is expected according to a specific hypothesis. A chi-square test is typically used with nominal or ordinal level data, thus instead of using means and variances, this test uses frequencies.

Correlation - A common statistical analysis, usually abbreviated as r , that measures the degree of relationship between two interval variables in a sample. The range of correlation is from -1.00 to $+1.00$. A correlation of -1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down. A zero correlation indicates there is no relationship between the variables. A correlation of $+1$ indicates a perfect positive correlation, meaning that both variables move in the same direction together.

Inferential statistics - Analyses conducted to determine the relationship between/among variables, generally with the goal of applying findings to the population the sample represents. Inferential statistics are also used to determine the probability that the identified relationships might have occurred as a result of chance.

IT-Test - A statistical test used to determine if the scores of two groups differ on a single variable.

Variable - Observable characteristics that vary among individuals.

Sample and sampling

Cluster sample – A sampling approach that can be used when a population is divided into groups (e.g., classes, grades, departments). Groups, rather than individuals, are the subject of sampling, such that all participants in the selected group(s) are part of the sample.

Convenience sample – Selecting participants who are easiest to access for participation in a study.

Population - The target group under investigation, as in all youth between the ages of 5 and 17 who live in Minnesota. The population is the entire set under consideration. Samples are drawn from populations.

Random sampling - Process used in research to draw a sample of a population strictly by chance, yielding no discernible pattern beyond chance. Random sample selection is used under the assumption that sufficiently large samples assigned randomly will exhibit a distribution comparable to that of the population from which the sample is drawn.

Sample – A sample is the subset of a population who are selected to participate in a study/evaluation. In studies that are meant to produce generalizable results, attempts are made to select a sample that is considered representative of the population to whom results will be transferred. Sample size is critical to generalizability as well - generally the larger the number in the sample, the higher the likelihood of a representative distribution of the population.

Stratified sample – A sampling strategy in which the population is divided into groups, then sample is drawn from each group.

Adapted from: Colorado State University's Writing at CSU website article, "Glossary of key terms" at <http://writing.colostate.edu/guides/research/glossary/>

Also includes information from the following sources:

Quizlet at <http://quizlet.com>

Mphekwane Mamahlodi's "What is the chi-square statistic?" at <http://cnx.org/content/m13487/1.2>

Significance/accuracy of findings

Confidence Interval - The range around a numeric statistical value obtained from a sample, within which the actual, corresponding value for the population is likely to fall, at a given level of probability.

Confidence Level - The specific probability of obtaining some result from a sample if it did not exist in the population as a whole, at or below which the relationship will be regarded as statistically significant.

Generalizability - The extent to which research findings and conclusions from a study conducted on a sample can be applied to the population at large.

Probability - The chance that a phenomenon has of occurring randomly. As a statistical measure, it shown as **p** (the "p" factor).

Reliability - The extent to which a measure, procedure or instrument yields the same result on repeated trials.

Validity - The degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure.

**Wilder
Research**

Information. Insight. Impact.

451 Lexington Parkway North
Saint Paul, Minnesota 55104
651-280-2700

www.wilderresearch.org



AMHERST H.
WILDER
FOUNDATION
ESTABLISHED 1906

Here for good.