Data Interpretation for Idaho Public Health Professionals

Glossary



- **Bias:** A systemic error that results in an incorrect or invalid estimate of the measure of association.
- Bimodal distribution: A distribution having two distinct statistical modes.
- Chance: An unexpected, random, or unpredictable event.
- **Confidence interval:** A range of values computed in such a way that it contains the estimated quantity a high proportion of the time.
- Consistency: Reliability or uniformity of successive results or events
- Correlation: A synonym for association or the relationship between variables.
- Data: Numerical information. Data is a plural term; the singular is datum.
- **Dependent variable:** A variable that may be predicted by or caused by one or more other variables called independent variables. For example, if it is believed that age influences the frequency of delinquent behavior, age is the independent variable and frequency of delinquent behavior is the dependent variable.
- **Descriptive statistics:** A statistic used to describe a set of cases upon which observations were made. Descriptive statistics include quantitative information such as measures of central tendency (mean, median, mode), confidence interval, and p-value.
- **Independent variable:** A variable that may predict or cause fluctuation in an dependent variable. For example, if it is believed that age influences the frequency of delinquent behavior, age is the independent variable and frequency of delinquent behavior is the dependent variable.
- Mean: The arithmetic mean is what is commonly called the average. The mean is the sum of all the scores divided by the number of scores.
- Median: The median is the middle of a distribution: half the scores are above the median and half are below the median. It is preferred when the distribution of cases is highly skewed since the median is not sensitive to outliers.
- **Mode:** The mode is the most frequently occurring score(or scores) in a distribution.



Data Interpretation for Idaho Public Health Professionals

Glossary



- **Person-years:** The total time of experience or exposure of a *group* of people who have been observed for varying periods of time.
- **Population:** An entire set of objects, observations, or scores that have something in common.
- **Outlier:** An instance that does not fit with other instances, or in other words, an instance that, compared to other members of a population, are at the extremes on relevant dimensions.
- Sample: A sample is a subset of a population.
- Sample size: The size of a sample. "N" is used to indicate the sample size. For example, if you have a sample of 23 people, you would write n=23.
- **Standard population:** A population used to allow comparisons over time and among different parts of the population. By convention the standard population is the US population in the year 2000.
- Statistical significance: The degree to which a value is greater or smaller than would be expected by chance. Typically, a relationship is considered statistically significant when the probability of obtaining that result by chance is less than 5% if there were, in fact, no relationship in the population.
- Variable: A variable is any measured characteristic or attribute that differs for different subjects.

