

| | | |
|---------------------------|--------------------------------------|----------------------|
| Data | Statistics | Population |
| Census | Sample | Correlation... |
| Voluntary Response Sample | Statistical & Practical Significance | Quantitative Data |
| Qualitative Data | Discrete Data | Continuous Data |
| Fewer vs Less | Ratio Measurement | Interval Measurement |
| Ordinal Measurement | Nominal Measurement | Big Data |
| Experiment | Experimental Units | Observational Study |

| | | |
|---|--|--|
| Collection of <i>all</i> measurements or data, e.g. data from group or object of study. | Science of collecting, organizing, analyzing, interpreting, and presenting data. | Collection of observations, e.g. measurements, colors, responses, etc.. |
| ... does not imply causation. | <i>Subcollection</i> of members selected from a population. | Collection of data from <i>every</i> member of a population. |
| Data consisting of numbers with meaning relative to the object measured. Also called numerical data. | Statistical significance is an event unlikely to occur by chance. Practical significance is whether an event makes a ‘real-life impact.’ | Where respondents decide whether to be included. Also called self-selected-sample. |
| Data where the data values are quantitative and any number between two possible data values is possible. The data values are not countable. | Data where the data values are quantitative and finite or countable. There are “gaps” between data values. | Data consisting of categories, names, labels, etc.. Also called categorical or attributive data. |
| Data which can be ordered and difference make sense. However, there is no natural zero. Example: Year, Temperature °F | Data which can be ordered, have a natural zero, and in which differences/ratios make sense. Example: heights, times | Fewer is used for describing amounts whereas less is for continuous amounts. For example, “fewer cans of soda” but “less distance traveled”. |
| Data sets so large and complex that their analysis defies traditional software tools. | Data which is organized by category. There is no ordering. Example: Yes/No, Coded Answers | Data which can be ordered but differences are meaningless. Example: Course Grade, Ratings |
| An experiment involving observation and measurement of characteristics but no attempt to modify the experimental units is made. | The individuals in an experiment. When there are people, we refer to them as subjects. | A process of applying a process or treatment and observing its effects. |

| | | |
|-------------------------|------------------------|----------------------------|
| (Double) Blinding | Random Sample | Simple Random Sample (SRS) |
| Confounding Variable | Systematic Sampling | Convenience Sampling |
| Stratified Sampling | Cluster Sampling | Multistage Sampling |
| Cross-Sectional Study | Retrospective Study | Prospective Study |
| Meta-analysis | Matched Pairs | Descriptive Statistics |
| Inferential Statistics | Parameter | Statistic |
| Characteristics of Data | Frequency Distribution | Lower Class Limit |

| | | |
|--|---|--|
| <p>A sample where each group of the same size is equally likely to be chosen.</p> | <p>A sample where each member of the population is equally likely to be chosen.</p> | <p>A study where the subject does not know if they are receiving treatment or not. Double blinding is where both the subject and experimenter do not know who is receiving treatment and who is not.</p> |
| <p>A sampling where data which is easily obtained is used.</p> | <p>A sampling where subjects are chosen at regular intervals.</p> | <p>A variable which affects the dependent and independent variable but is not included in the study. Also called a hidden variable or lurking variable.</p> |
| <p>A survey where different survey methods are used at various stages.</p> | <p>Where a population is broken into groups (often with different characteristics) and a sample (of all) individuals is selected from the groups.</p> | <p>Where a population is broken into at least two groups (with similar characteristics) and a sample is chosen from each group.</p> |
| <p>Where data is collected in the future from groups with a common factor. Also called a longitudinal or cohort study.</p> | <p>Where data is collected from a past period. Also called a case-control study.</p> | <p>A study where data is observed, measured, and collected at one point in time, not over a period of time.</p> |
| <p>Describes data, e.g. how one presents data or interesting characteristics: mean, median, min/max, standard deviation, skew, kurtosis, etc..</p> | <p>A design where subjects are matched into groups or pairs having similar characteristics.</p> | <p>A method where one combines data and/or results from multiple independent studies.</p> |
| <p>A numerical measurement used in describing some characteristic of a <i>sample</i>. “Sample statistics”</p> | <p>A numerical measurement used to describe some characteristic of a <i>population</i>. “Population parameter”</p> | <p>Draws inferences from data, e.g. z-statistic, t-statistic, χ^2, F-ratio, etc..</p> |
| <p>The smallest number in each class.</p> | <p>Shows how data are partitioned among several categories (or classes) by listing the categories along with the number (frequency) of data values in each of them.</p> | <p>center, variation, distribution, outliers, time</p> |

| | | |
|------------------------------|--------------------------------|-----------------------------------|
| Upper Class Limit | Class Boundaries | Class Midpoints |
| Class Width | Relative Frequency | Cumulative Frequency Distribution |
| Skewed | Correlation | Linear Correlation |
| Correlation Coefficient, r | (Linear Regression) p -value | Mean |
| Median | Mode | Midrange |

| | | |
|---|--|---|
| <p>The middle value of each class.</p> | <p>The “middle” number between classes, i.e. average of the class limits.</p> | <p>The largest number in each class.</p> |
| <p>A frequency distribution in which the frequency in each class is the sum of the frequencies for the previous classes and that class.</p> | <p>Relative Frequency $= \frac{\text{frequency for class}}{\text{sum of frequencies}}$</p> | <p>The length of each class, i.e. the difference between the upper and lower class boundaries.</p> |
| <p>When two variables are correlated and the plotted pairs of points result in a pattern that can be approximated by a straight line.</p> | <p>When two more more variables are somehow associated with one another.</p> | <p>When a distribution is not symmetric and extends more to one side than the other.</p> |
| <p>The ‘average’ value of a data set: $\frac{\sum x_i}{n}$</p> | <p>The probability that getting a paired sample with a linear correlation coefficient r that is at least as extreme as the one obtained from the paired sample.</p> | <p>Measures the strength and direction of a linear relationship. r^2 measures the “percent linearity.”</p> |
| <p>A value midway between the minimum and maximum value of a data set: $\frac{\text{min} + \text{max}}{2}$</p> | <p>The value which repeats most in a data set. A set may have no mode, one mode, two modes (bimodal), ore more (multimodal).</p> | <p>The ‘middle’ of a data set.</p> |