

MATHEMATICS 201-510-LW

Business Statistics

Martin Huard

Fall 2008

Definitions

Statistics is the study of how to collect, organize, analyze data and interpret numerical information from data.

Individuals are the people or objects included in the study.

A **variable** is the characteristic of the individual to be measured or observed.

Variables are either **quantitative** (has a numerical value on which operations such as additions make sense) or **qualitative** (refers to a category or group) in nature.

In **population data**, the variable is from every individual of interest.

In **sample data**, the variable is from only some of the individuals of interest. A sample is a subset of a population.

Descriptive statistics involves methods of organizing, picturing, and summarizing information from samples or populations.

Inferential statistics involves methods of using information from a sample to draw conclusions regarding the population.

Levels of Measurement

Nominal level: Data usually consisting of names. There is no logical order.

Ordinal level: There is an order, but actual differences between data value are meaningless.

Interval level: The measurement is a numerical value, where differences between data values can be computed. However, ratios of values are meaningless.

Ratio level: It is like the interval level, but it includes an inherent zero. Both differences and ratios are meaningful at this level.

A **(simple) random sample** of n measurements from a population is one that is selected in a way such that

(i) every element in the population has an equal chance of being selected

(ii) every sample of size n from the population has an equal chance of being selected.

A random number table or a random number generator (such as Excel) is usually used to pick the sample.

Other Sampling methods

Stratified sampling: A strata is a group or class of a population that share a common characteristic, for example, men and women. In stratified sampling, the population is divided into stratum, then a (simple) random sample is taken in each stratum, and the information is carefully adjusted.

Systematic sampling: Here we assume that there is an order in the elements of the population, and, starting at some point, we proceed to select every k^{th} element for our sample.

Cluster sampling: we divide the population into areas or sections (clusters), then randomly select sections and include all members in the sample.

Convenience sampling: Uses data that are conveniently and readily obtained. Note: convenience sampling runs the risk of being severely biased.

Guidelines for planning a statistical study

1. Identify the individuals or objects of interest.
2. Specify the variables as well as protocols of procedures for taking measurements or making observations.
3. Determine if you will use the entire population or a sample. If using a sample, decide on a viable sampling method.
4. Collect the data.
5. Use appropriate descriptive statistics methods and make decision using appropriate inferential statistics methods.
6. Finally note any concerns you might have about your data collection methods and list any recommendations for future studies.

Census: measurements or observations are taken from the entire population.

Sampling: measurements or observations are taken from a representative part of the population.

Simulation: a numerical facsimile of real-world phenomena. Often uses computers.

Observation study: Observations and measurements of individuals are being conducted in a way that doesn't change the response or the variable being measured.

Experiments: A study in which some treatment is deliberately imposed on units or subjects in order to observe a given response.

Survey: Data about people is gathered by asking them questions. Note: there are many potential pitfalls.

Hidden Bias: You are not measuring what you hoped to measure.