The basics of statistics

(STATS201.stat 202 10: Experimental design and descriptive statistics)

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 11, 2012

Sample and population

Population

The **population** of a study is the set of all subjects that inference should be made about.

Sample

A **sample** is a set of subjects that are sampled from a given population.

- Every sample can only be sampled from one population.
- Different samples can be sampled from the same population.

Categorical and numerical variables

Variable

A **Variable** is a certain property that is noted or measured on the subjects in the sample.

Categorical variables

Categorical variables do not have numerical values but, as the name suggests, indicate which category the subject belongs to.

Numerical variables

Numerical variables have numerical values that are measured in some units.

Continuous and discrete variables

Continuous variables

When a numerical variable can have any numerical value on some interval it is referred to as **continuous**. Only numerical variables can be continuous.

Discrete variables

If variables are not continuous they are referred to as **discrete**. All categorical variables are discrete and some numerical variables.

Exploratory and response variables

For every subject, the value of an **explanatory variable** influences the value that its **response variable** will obtain. A response variable can be influenced by several explanatory variables.

Randomness

We apply statistics because our measurements are influenced by some randomness:

- We measure only a sample of the whole population.
- The phenomena to be measured are random by nature.

This property is described with the concept **random variable**.

A random variable describes the outcome of a variable before it is measured.