Statistics: 101

Intro

- Statistics: used by scientists to convert data into useful information
- This is a process, divided in three main stages:

1. Producing Data
2. Exploratory Data Analysis
3. Interpreting Data

Definitions

- Population: entire group of interests, people objects, plants, etc.

- Sample: sub-set of the population.

- Data: information about individuals organized in variables

- Observation: particular unit of observation within the sample (e.g., person object)

- Variable: particular characteristics of the unit of observation

- Data set: set of data of individual observations and variables

- Codebooks: offer complete information regarding the data set

- Raw data: lists of numbers and labels, not very informative, lacks context

Types of Variables

- Quantitative: take numerical numbers, represent a measurement (eg, age, weight).

- Categorical (Qualitative): take category or label values, place an observation in a group (e.g., race, gender)

- Also important:

- Dummy codes: numbers used to represent categories (or labels), they have no arithmetic meaning.

- Unique identifier: variable meant to distinctively define each unit of observation (e.g., serial numbers for objects, social security numbers for persons)