

Study guide and primer on statistics concepts

Descriptive statistics

You must know definitions of

1. **Population, Parameter, Sample, Statistic**
2. **Types of Data (Variables)**
3. **Measure of Center, Variability and Relative Variability (Information and Insights)**
4. **Graphical and Numerical Summary**
5. **Frequency Distribution, Histograms, Boxplots, Empirical rule**

You must also know how to 'compute' or 'plot' some of these given data. There are many references online that will help you to do this. The key thing is to know the terminology.

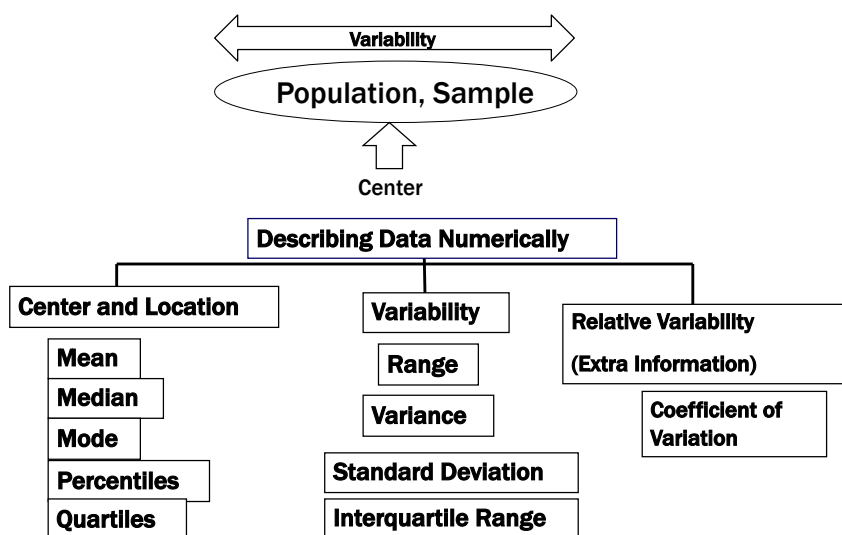
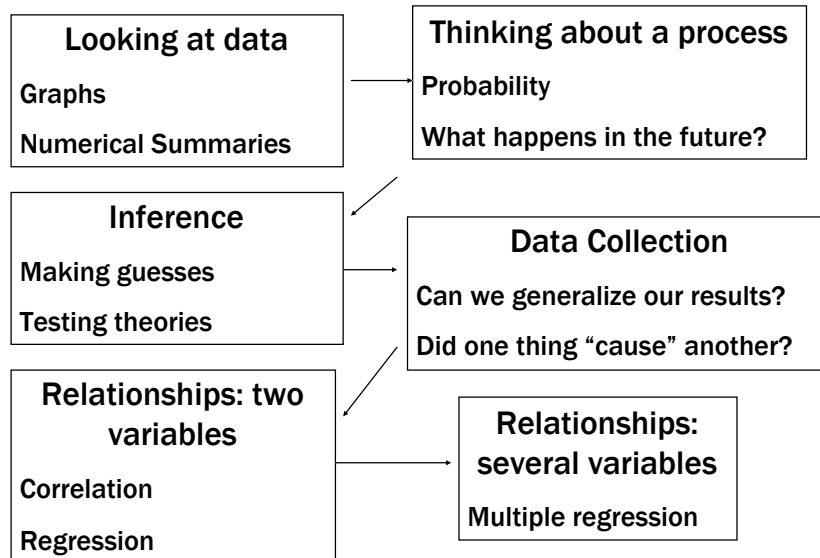
Let's review some basic statistical concepts

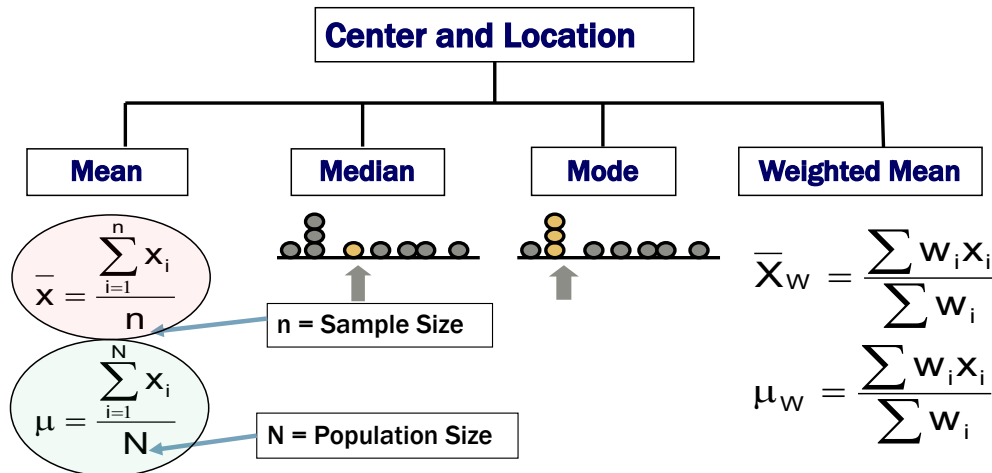
- **Population:** The *total* collection of elements of interest in a given study.
 - *All students in USC*
 - *all possible employees of a company*
- **Variable:** A measurable characteristic that takes on different values.
 - **Parameter:** A number that provides a quantitative description of some characteristic of a *population*. Generally, the average or proportion
 - *proportion of people who rent*
 - *average GPA*
- **Sample:** A subset of a population.
 - *1,000 individuals in a health survey*
 - *5 students selected at random*
 - **Statistic:** A numerical quantity that provides information concerning some characteristic of a *sample*. Generally, the average or proportion
 - *proportion of home-owners who are seeking assistance for health reasons in the sample of 1,000*
 - *average age of the students in the sample of 5*
- **Observation:** A recorded value of a variable for a particular individual.
 - *this guy is 5' 9" tall*
 - *this person is a Republican*
- **Data:** A set of observations (i.e. the numerical values recorded for all individuals in a sample).
- **Variable:** A measurable characteristic that takes on different values.
 - *Quantitative variable: differs in amount (e.g. yearly pay)*

– Qualitative variable: differs in kind (e.g. party affiliation)

- Statistical inference: The use of *sample* information to infer something about *population*.

Relationship between some key statistical concepts





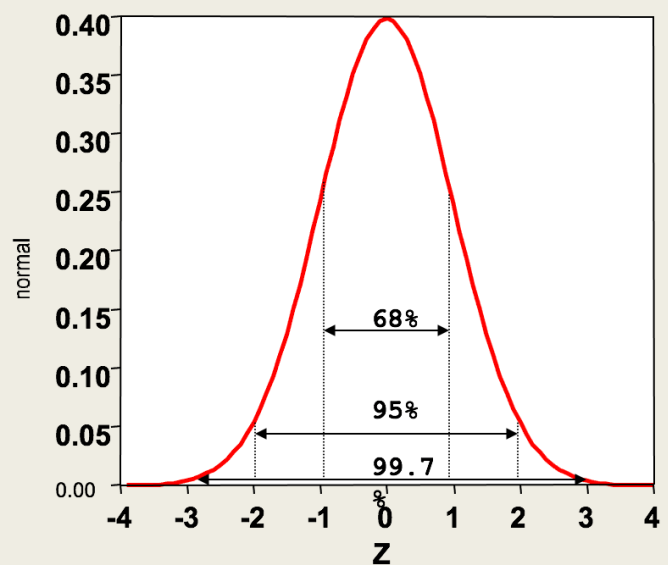
Empirical Rule

If the normal curve fits well then (approximately):

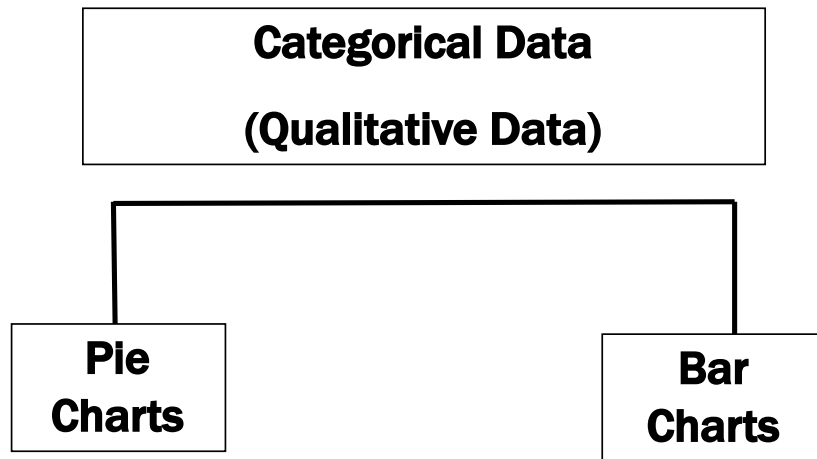
68% of the data is within 1 SD (Standard Deviation) of the mean.

95% within 2 SD.

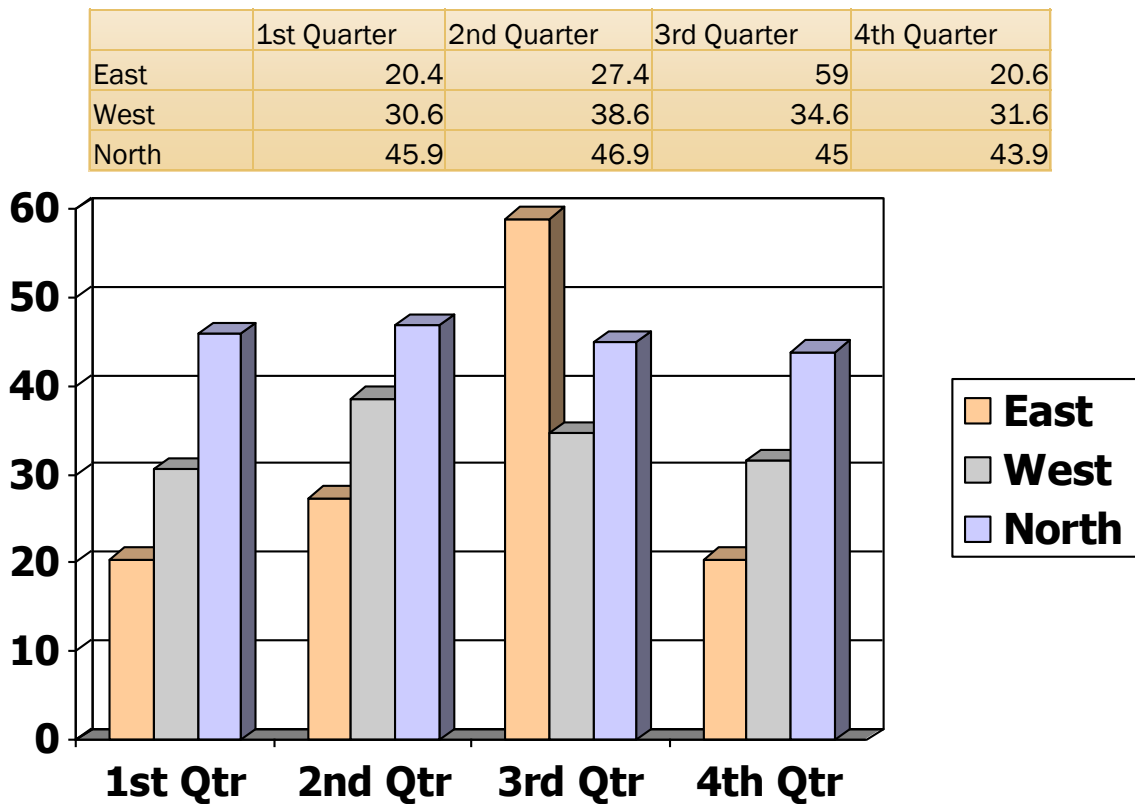
99.7% within 3 SD.



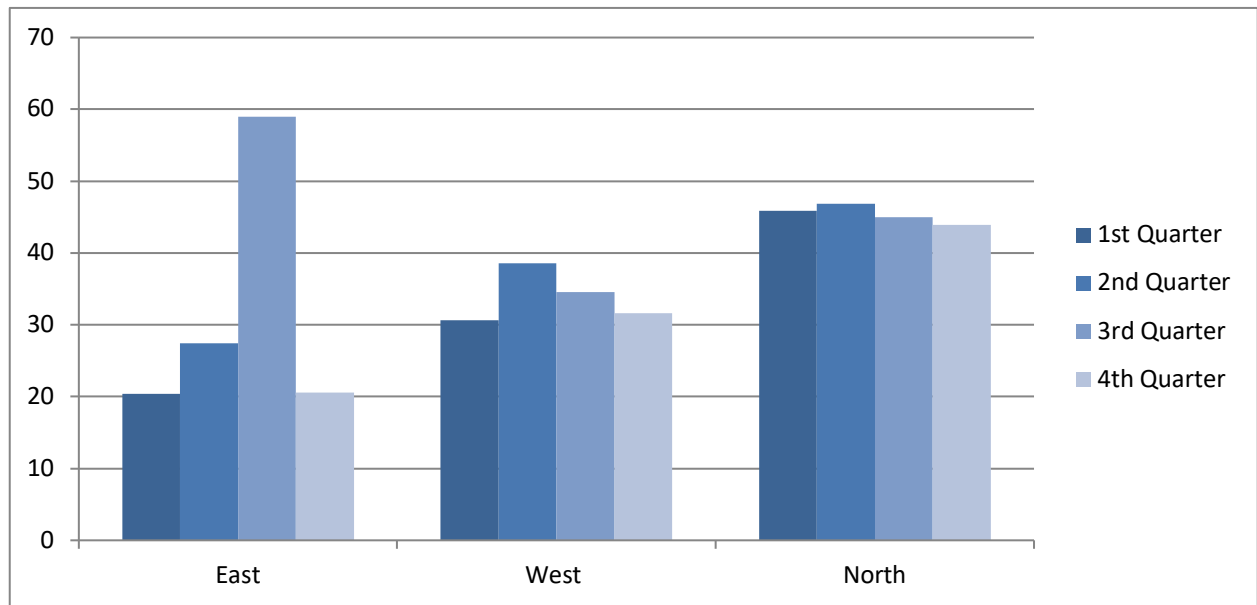
Graphical categorical data



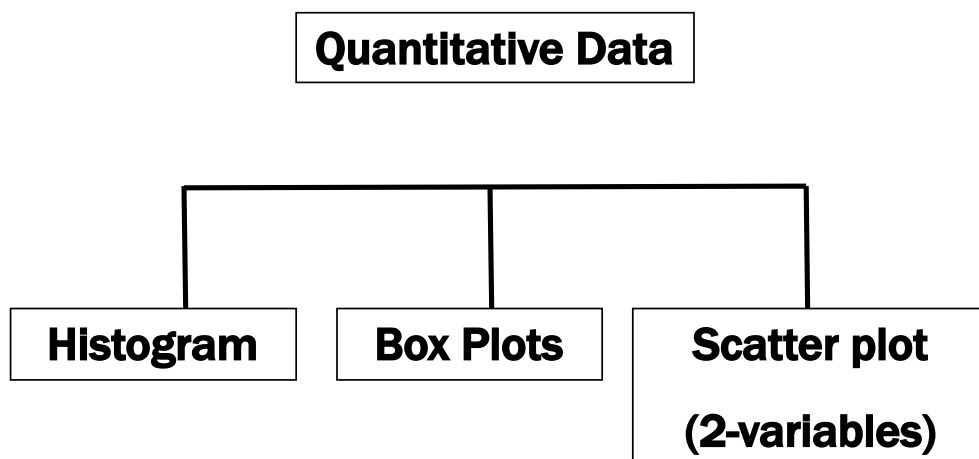
What if we have two categorical variables?



Alternate view (which is better?)



Graphing Quantitative data



Data types

Nominal : simply classification.

- party affiliation
- brand preference

Ordinal : values are ordered (large numbers indicate greater amounts).

- rate this coffee (1 = "Great",, 5 = "Awful")
- on a scale of 1 to 10, what do you think of President Reagan?