Statistics

Goal: use data to draw insights about population.

Population ⇝ well defined collection of objects ⇝ interested in studying.

Example

⇝ ① people in the USA
② Cars in Maryland
③ Televisions manufactured in a factory
④ people who took the vaccine → safety of vaccine.
⑤ Senior Citizens living in NJ → how environment affect senior.
⑥ Stat 400 students in UMD → which students did better, avg GPA.

Q/C: age, pre-existing conditions, weight, height
C: sex
Q: level of physical activity, retirement benefits.
C: location, major of study, ethnicity.
Q: GPA

Variable ⇝ Characteristic of an object in the population

Quantitative
⇝ measurements, counts, numeric values
⇝ have units

Qualitative/Categorical
⇝ Qualitative properties like color, sex, race, nationality, etc
⇝ Can put into categories.

Goal Restated:

Draw insights/inferences of properties of variables at the level of the population

Goal Restated:

Draw insights/inferences of properties of variables at the level of the population

↓

Problems

① Population might be very large / intractable.

② Too expensive to access all objects in population

③ All objects in the population might not be accessible.

↓

Ex: studying certain fungi in the Amazon rainforest.

Solution:

Work with a subset of the population

↓

sample

Generally:

size of sample $\ll$ size of population

Advantages →
- have access to all objects in the sample
- relatively small set of objects
- ease of calculations

Work with a subset of the population

⇓

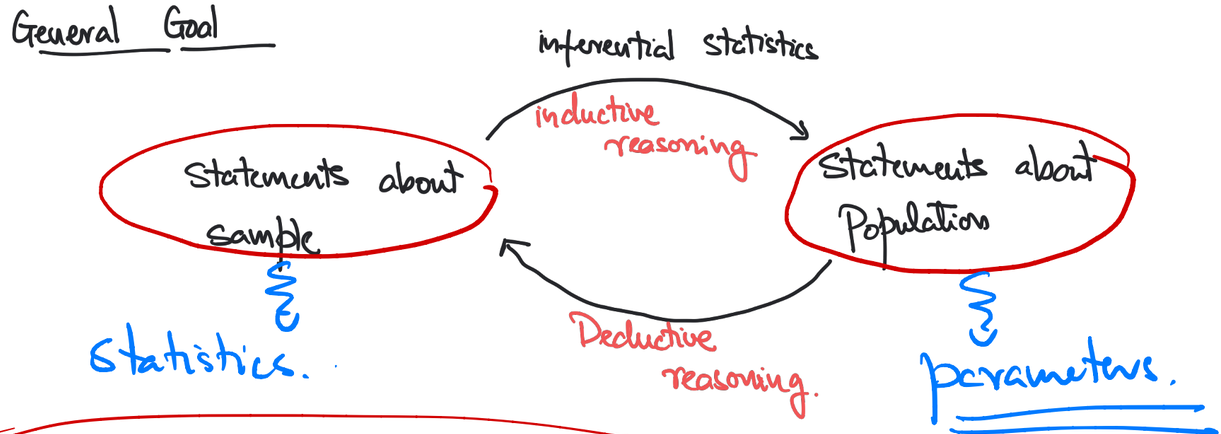
sample

Generally:

size of sample $\ll$ size of population

Advantages →
- have access to all objects in the sample
- relatively small set of objects
- ease of calculations

General Goal



Note: There is no free lunch!

The price of working with a subset of the population

⇓

additional care in choosing the sample.

$$\int_{-\infty}^{\infty} f(x)dx = C < \infty$$

$$\int_{-\infty}^{0} f(x)dx < \infty \quad \text{and} \quad \int_{0}^{\infty} f(x)dx < \infty$$

$$\lim_{c \to -\infty} \int_{c}^{0} f(x)dx \qquad \lim_{c \to \infty} \int_{0}^{c} f(x)dx$$

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} f\left(\frac{x-b}{a}\right)dx$$

location-scale families

$$\lim_{c \to -\infty} \int_{c}^{0} f\left(\frac{x-b}{a}\right)dx \qquad \lim_{c \to \infty} \int_{0}^{c} f\left(\frac{x-b}{a}\right)dx.$$

Hint: apply change of variables.