Basics of Statistical Inference: Overview of Sampling

Mainak Thakur

25.08.2020

Population

- “Every statistical enquiry is designed to gather information about some aggregate of individuals— individual objects or beings —rather than about the individual themselves. In statistical language, such an aggregate is called a population or universe.”
- Enquiry related to human population of the world, population of tigers in the world, population of all the trains in India, all the employees of Walmart, population of all the products produced in an industry or a machine.
- Boundary: time and space
- Finite and infinite

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Sample

- A small number of members of the population (The data in hand of the investigator)
- Used to investigate behaviour or characteristics of population concerned
- Example: An NGO wants to calculate for a year the average income per family for all families residing in Andhra Pradesh. To accurately know this figure, the NGO needs to know the income of several lakhs of families living in the state.
- Time and money
- Choose a sample of 1000 or 10000 wisely and make conclusions based on the characteristics of the sample

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Sampling

- The n families to be appeared in the sample: 1st n families, last n families in National Register, or every 100th family starting from the 1st etc.
- Defective methods, unrepresentative samples, several drawbacks
- Probability sampling: each member of the population gets a definite probability of being included in the sample
- Commonly used: simple random sampling (SRS) or random sampling, each member of the population has the same probability of being included in the sample

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

SRSWR and SRSWOR

- SRSWR: the n samples are drawn from the population one by one, after each drawing the individual selected are returned to the population. At each drawing, each of the N members of the population have same probability 1/N of being selected.

- SRSWOR: The n members of the sample are drawn one by one but the individuals are not returned to the population. At each stage every remaining unit of the population is provided the same probability of being included in the sample. At the rth drawing (r=1,2,3,…,n) , each member of the population get probability 1/ (N-r+1) of being actually selected.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Parameter and Statistic

- Statistical investigations: in general, aim is to understand characteristics of the individuals of the population. Sampling provides understanding of characteristics of the members of the sample.
- Assume only one character you are interested in, x
- x_i, the value of x for the ith member of the sample, then $x_1, x_2, \ldots, x_n$ are the sample observations
- Ultimately want to know the values of different measures of the population distribution of x, like its mean, s.d etc. The measure of this type, calculated on the basis of population values of x, is called a parameter.
- A corresponding measure calculated on the basis of sample values is called a statistic.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Parameter and Statistic

- Population members included in different samples from the same population may be different, the value of the statistic varies from one sample to another.
- Sampling distribution of the statistic
- Population parameter is generally considered as constant
- The nature of the sampling distribution of a statistic can be deduced theoretically, provided the nature of the population is given
- Sampling distribution has mean, s.d. and higher order moments

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Statistical Inference

- Features of sample differ from the feature of the population
- What can be said about the properties of the population from a knowledge of the properties of the sample?
- Sampling theory addresses this question
- The process of going from the known sample to the unknown population has been called statistical inference

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Estimation and Testing of Hypothesis

- Some completely unknown feature of interest of population may be guessed on the basis of a random sample from the population.
- Problem of Estimation.
- Some information about the feature may be available and he may want to check whether the information is tenable in the light of the random sample taken from the population.
- Problem of testing of hypothesis.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Before and after data collection

- Variable of interest for Population: $X \sim N(\mu, \sigma^2)$
- Drawing n samples: Before data collection

$$\underset{n\times 1}{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{bmatrix}$$

After Data collection

$$x_{n\times 1} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta
nptel, https://www.youtube.com/watch?v=ZAYt7_bf6vM&list=PLbMVogVj5nJRt-ZxRG1KRjxNoy7J_IaW2&index=3