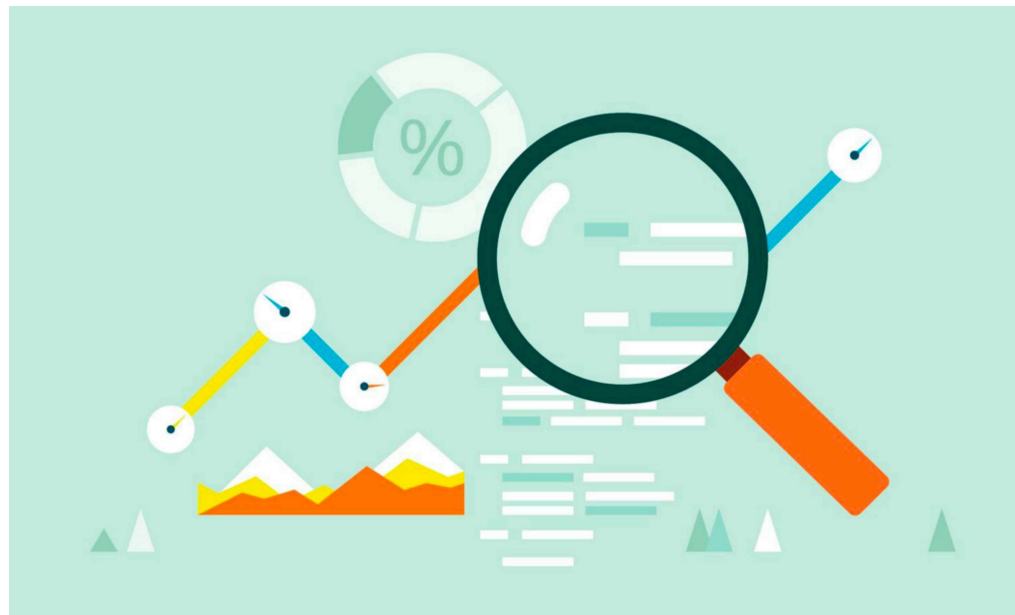


Data Science





Statistics

Statistics



- Statistics is a collection of tools and methods that you can use to derive meaningful insights from data by performing mathematical computations on it.

Statistics in Data Science

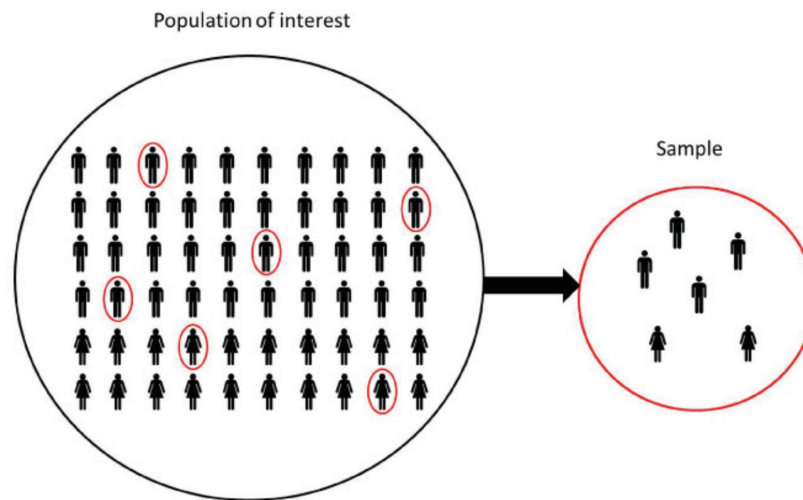
- “Data Scientist is a person who is better at statistics than any programmer and better at programming than any statistician.”



Population & Sample

Population & Sample

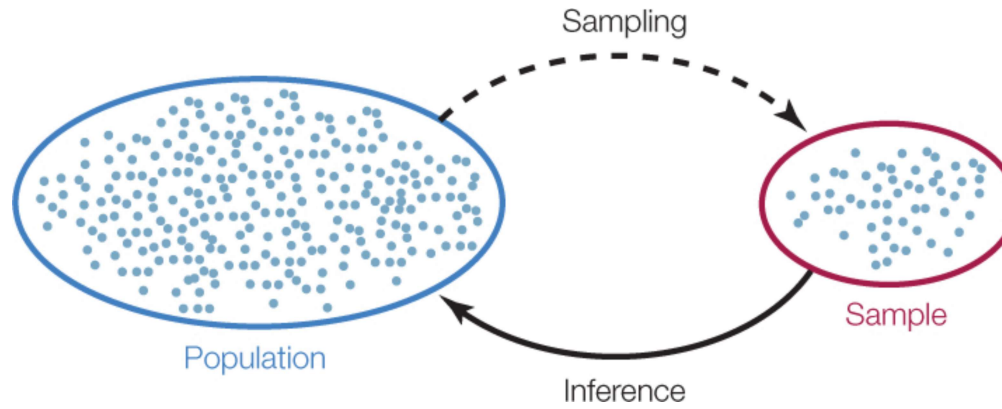
- A set of all individuals relevant to a particular statistical question is called a population.
- A smaller group selected from a population is called a sample.



Parameter & Statistic

- Parameter is descriptive measure of population and statistic is descriptive measure of sample.

Sampling



Sampling Methods

- Probability sampling or Random Sampling
- Non-probability sampling

Probability sampling

- Simple Random Sampling
- Stratified sampling
- Cluster Sampling
- Systematic sampling
- Multi stage Sampling



Simple Random Sampling

Simple Random Sampling



- Every unit of the population has an equal chance of being selected

Simple Random Sampling

- Sampling without replacement
- Sampling with replacement

Simple Random Sampling



- Sampling Error



Stratified Sampling

Stratified Sampling

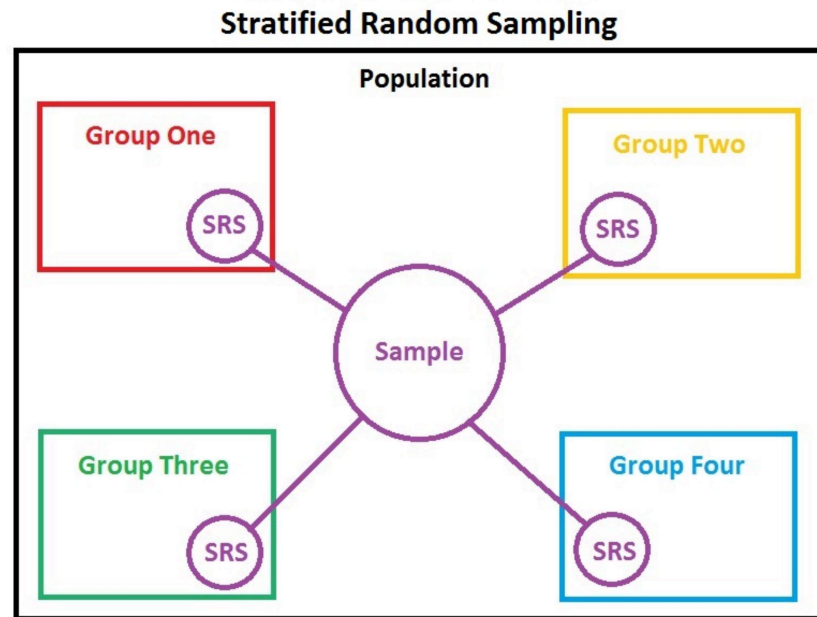


Image Source - https://en.wikipedia.org/wiki/Stratified_sampling

Stratified Sampling

- Subsets of the data sets or population are created based on a common factor, and samples are randomly collected from each subgroup.
- Should be used when we want some members from every group

Examples - Stratified Sampling



- Music taste of students
- Average Salary
- Average package offered



Cluster Sampling

Cluster Sampling

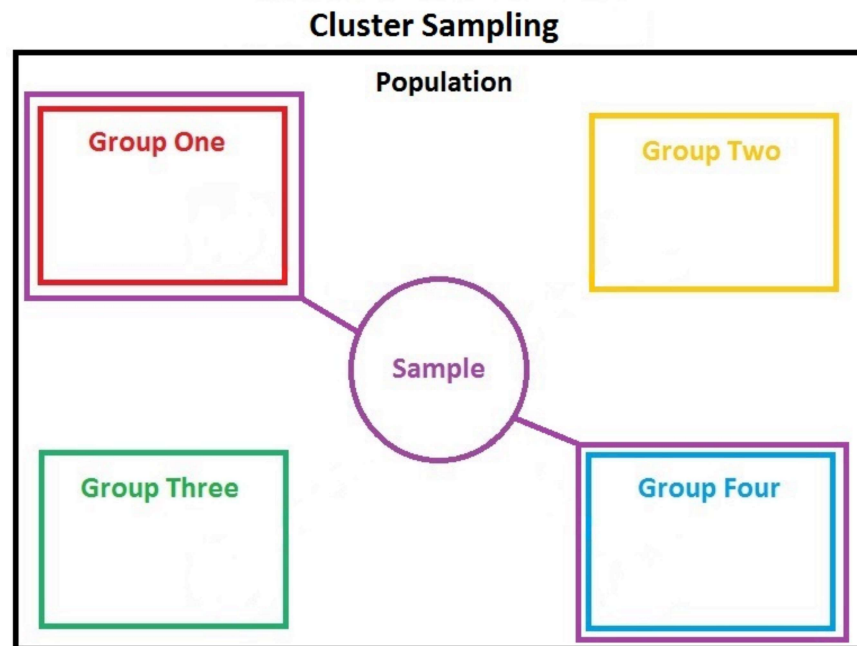


Image Source - https://en.wikipedia.org/wiki/Cluster_sampling

Cluster Sampling

- Single Stage Cluster Sampling
- Two Stage Cluster Sampling



Systematic Sampling

Systematic Sampling

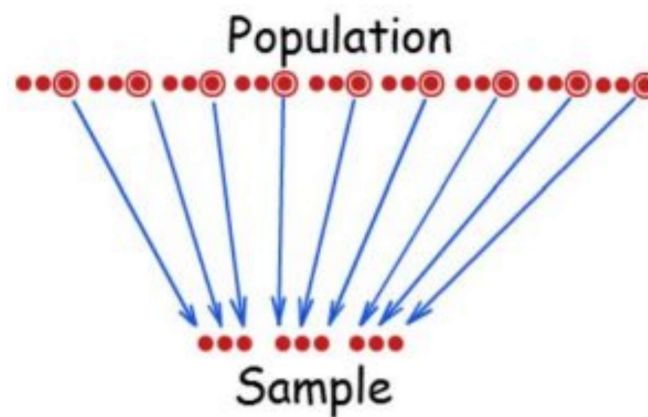


Image Source - <https://stepupanalytics.com/wp-content/uploads/2018/08/5-3-300x190.jpg>

Probability sampling

- Simple Random Sampling
- Stratified sampling
- Cluster Sampling
- Systematic sampling
- Multi stage Sampling