



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

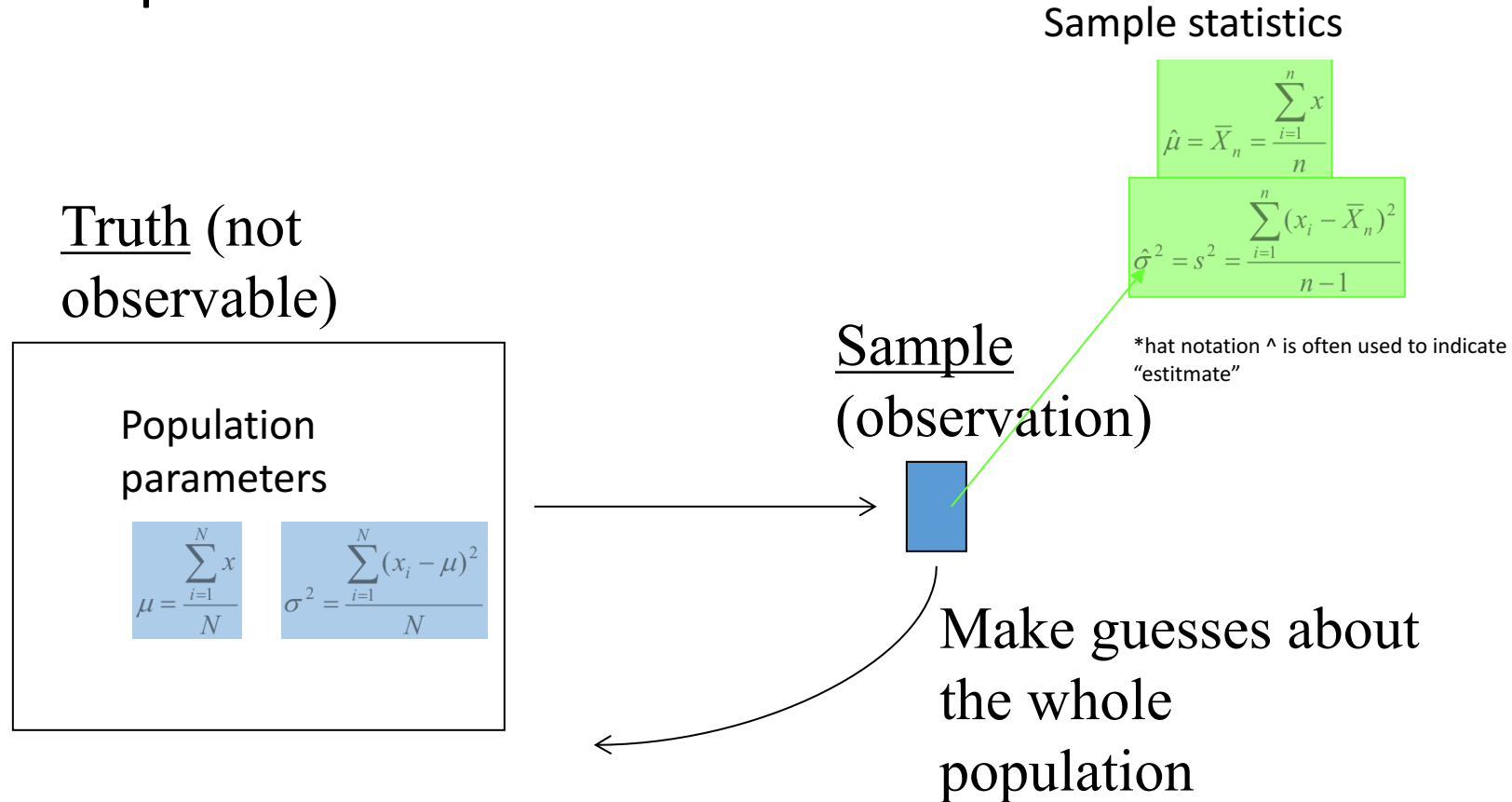
CS639: Data Management for Data Science

Lecture 14: Sampling

Theodoros Rekatsinas

Statistical Inference

- Statistical inference: The process of making guesses about the truth from sample data.



Introduction



Population



Sample

- Probability sampling
- Non-probability sampling

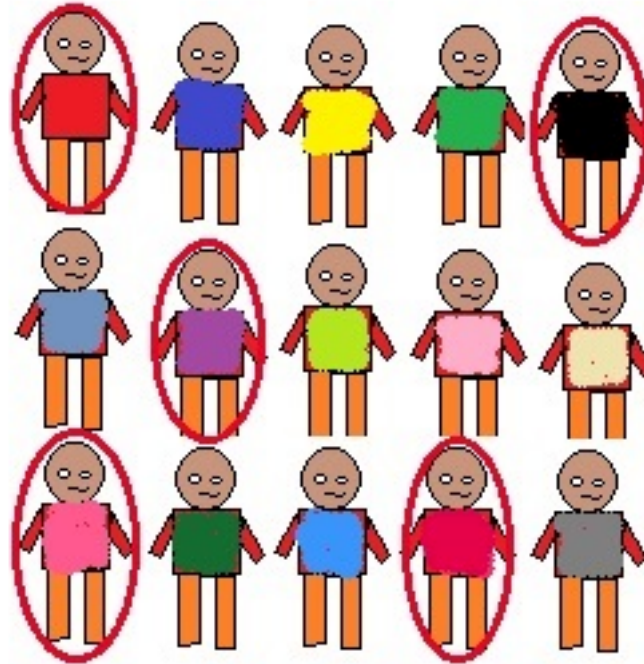
Probability sampling

- We use randomization to select elements from our population

Examples

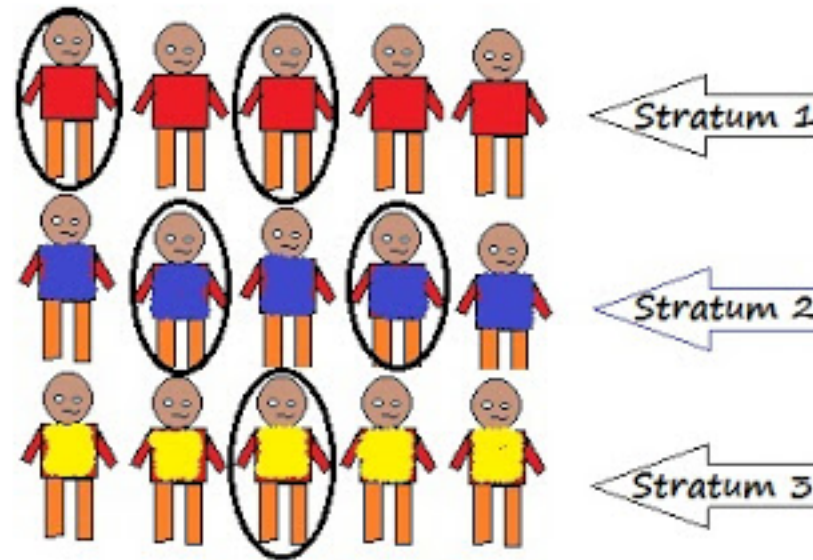
- Simple Random Sampling
- Stratified sampling
- Systematic sampling
- Cluster sampling
- Multi-stage sampling

Random sampling



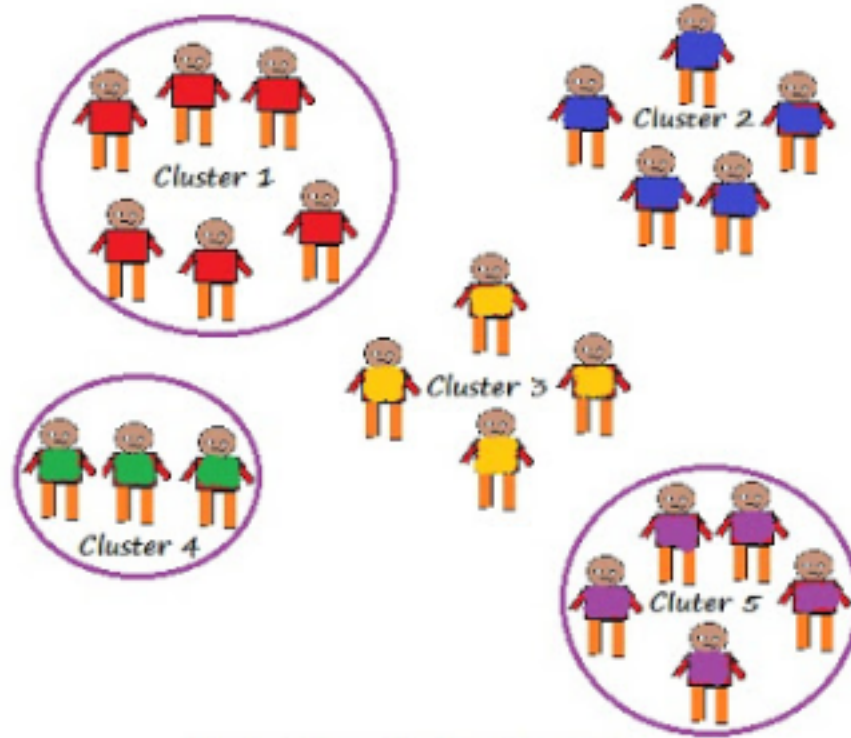
Random selection of 20 students from class of 50 student. Each student has equal chance of getting selected. Here probability of selection is $1/50$

Stratified sampling



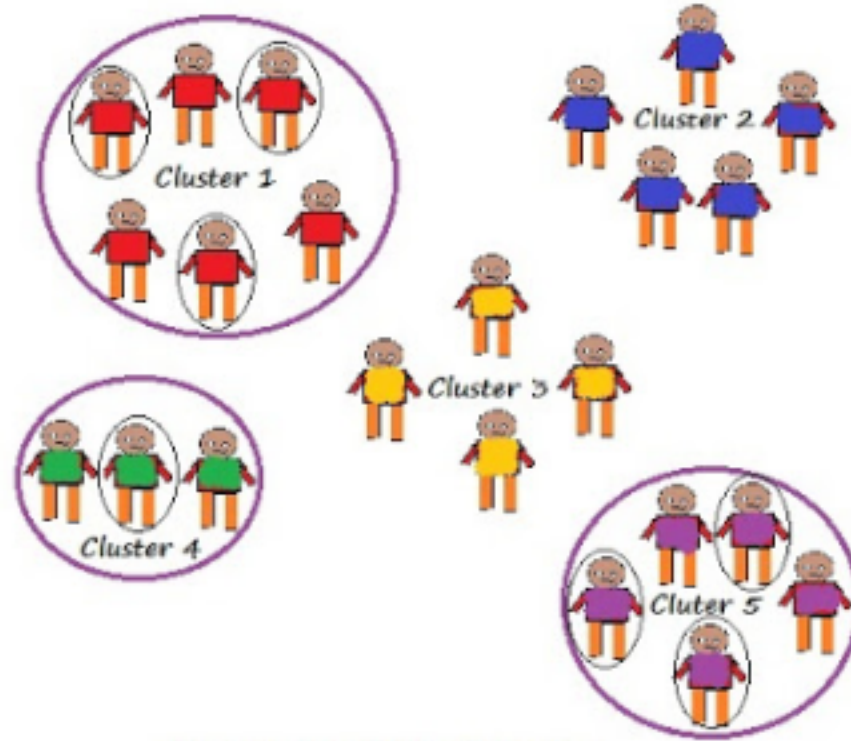
Divide the elements of the population into small subgroups (strata) based on the similarity in such a way that the elements within the group are homogeneous and heterogeneous among the other subgroups formed. And then the elements are randomly selected from each of these strata.

Single Stage Cluster sampling



Our entire population is divided into clusters or sections and then the clusters are randomly selected.

Two Stage Cluster sampling



First we randomly select clusters and then from those selected clusters we randomly select elements for sampling

Non-Probability sampling

- We use randomization to select elements from our population

Examples

- Convenience sampling
- Quota sampling
- Snowball sampling

Convenience sampling

The samples are selected based on the availability. This method is used when the availability of sample is rare and also costly. So based on the convenience samples are selected.

For example: Researchers prefer this during the initial stages of survey research, as it's quick and easy to deliver results.

Quota sampling

This type of sampling depends of some pre-set standard. It selects the representative sample from the population. Proportion of characteristics/ trait in sample should be same as population. Elements are selected until exact proportions of certain types of data is obtained or sufficient data in different categories is collected.

For example: If our population has 45% females and 55% males then our sample should reflect the same percentage of males and females.

Snowball sampling

This technique is used in the situations where the population is completely unknown and rare. We will take the help from the first element which we select for the population and use that element to recommend other elements who will fit the description of the sample needed.

