

Scientific Methodology in Computer Science

MO430A

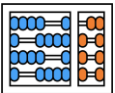
Prof. Dr. Bruno B. P. Cafeo

Institute of Computing
University of Campinas



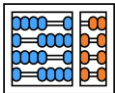
Agenda

- Type of data
 - Quantitative
 - Qualitative
- Population
- Classification of population
- Sampling



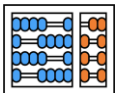
Definition

Data collection is the process of gathering data on certain variables in a structured and controlled manner, allowing one to analyse the data collected to answer relevant questions and assess consequences.



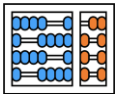
Type of data

Type of Data	Collecting methods	Analysis methods
Quantitative	Measuring and counting are used to collect quantitative data, with data stored in numerical form. Measurements can be made manually or using "sensors".	Quantitative data is analysed using statistical analysis. In some applications, the data may be analysed using more advanced techniques such as machine learning and artificial intelligence.
Qualitative	Interviews and observation are used to gather qualitative data. Data is usually descriptive, meaning there is no set range that results are limited to.	Qualitative data is studied by organizing it into meaningful groups or themes, which can then be further analysed. The data collected is non-numerical.



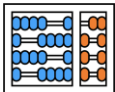
Qualitative data

Qualitative data is information defined using descriptive language, allowing open-ended, detailed answers that are non-numerical in nature.



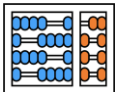
Qualitative data collection methods

- **One-on-One Interviews:** On a one-to-one basis, the interviewer or researcher obtains data from the interviewee, by asking them pre-prepared questions. These questions are designed to help the interviewer draw out information useful for making relevant conclusions.
- **Focus groups:** This type of research is done in a conversation format with a group of people. The group is generally restricted to 6-10 persons, and the debate is moderated by a moderator. The members of a group may have something in common depending on how the data is to be going to be sorted. A researcher doing a study on basketball players, for example, may select a group of people who are basketball players and a group of those who are not. They can then compare the data from the two groups.



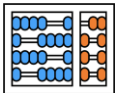
Qualitative data collection methods

- **Record keeping:** This strategy makes use of previously existing data and credible sources of information. This information can be utilized in future studies. It's like going to the library. There, one can go through books and other reference materials to gather essential facts for the research.
- **Case studies:** An in-depth investigation of case studies is used to acquire data in this strategy. This method can be used to evaluate both basic and complicated issues. The strength of this strategy is how well it draws conclusions using a combination of one or more qualitative data gathering methods. A case study could yield quantitative as well as qualitative data.



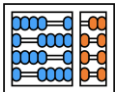
Quantitative data

Quantitative data refers to information that is expressed in numerical terms or as quantitative measurements. It represents data that can be counted or measured and is typically associated with a specific quantity.



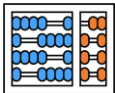
Quantitative data collection methods

- **Survey/Questionnaire:** In order to obtain data from a group or a large number of people, surveys or questionnaires are often used. Both quantitative and qualitative research can be collected from surveys/questionnaires. By controlling the questions people are asked data can be gathered from humans while restricting the range of answers, which makes the subsequent processing and analysing of data simpler.
- **Experiments:** Many hypotheses are tested through means of an experiment. This would involve carefully designing an investigation so that the variable(s) being measured are isolated from uncontrolled factors that could influence the result, and recording objective data about the experimental results using appropriately calibrated sensors and measuring equipment.



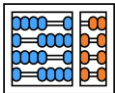
Planning a Research

- To conduct research, it is essential to select the population of interest.
- The population can consist of various types of objects, depending on the area of study.



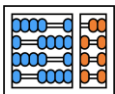
Population

- Population is defined as the set of objects that share at least one common characteristic.
- Population can be classified as finite or infinite and as the target population or the external population, depending on the criteria.



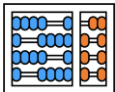
Classification of Population

- Examples of **finite** populations include students of quantitative methods or users of an application. **Infinite** populations include births in a city or machine production.
- The **target population** is the focus of the research, while the **external population** is the group to which the results are intended to be generalized.
- **Samples** are smaller subsets of the population that make research more feasible.



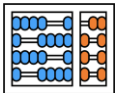
Sampling in Statistics

- The field of Sampling in Statistics deals with the composition of samples and employs various techniques and procedures.
- Types:
 - Simple random sampling
 - Systematic random sampling
 - Stratified sampling
 - Cluster sampling
 - Convenience sampling
 - Self-selection sampling
 - Purposive sampling
 - Snowball sampling
 - Quota sampling
 - Sample size calculation



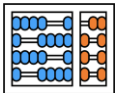
Simple random sampling

- Simple random sampling is the most elementary process in sampling. It is based on the principle that every member of a population has an equal chance of being included in the sample.
- To conduct this type of procedure, each participant in the population is assigned a number, which is drawn in a process sometimes referred to as a lottery. The sample is then comprised of the selected participants.



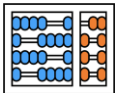
Systematic random sampling

- It is a variation of simple random sampling.
- After identifying the participants, a specific criterion is chosen (e.g., every 5th) and the selection follows this pattern.



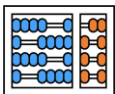
Stratified sampling

- In this type of sampling, the population is divided into subpopulations based on shared characteristics, known as strata.
- Each participant is then assigned an identification within their stratum.
- Simple random sampling is performed within each stratum.



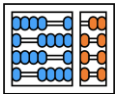
Cluster sampling

- In this type of sampling, the population is naturally located in clusters.
- These clusters can be streets, neighborhoods, or businesses, for example, and are assumed to be heterogeneous. The clusters are assigned identifications, which are then randomly selected. All participants in the selected clusters must be accessed.



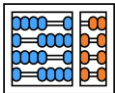
Convenience sampling

- In this type of sampling, the sample is taken from the participants to whom the researcher has the greatest access.



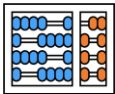
Self-selection sampling

- In this type of sampling, participants voluntarily request to take part in the research.
- This tends to occur in studies where data collection is conducted online (e.g., Google Survey or SurveyMonkey) and in research on new medications



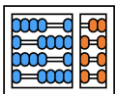
Purposive sampling

- Purposive sampling is a type of sampling in which the researcher deliberately selects the individuals or items to be included in the sample.
- It is frequently used in various computer science studies, such as usability testing, software validation, and expert evaluations.



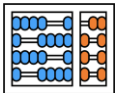
Snowball sampling

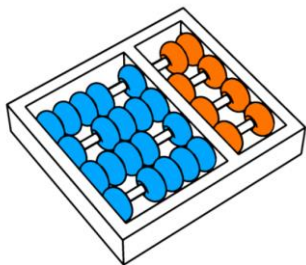
- In snowball sampling, the researcher identifies an initial participant of interest within a computer programming community who, in turn, refers other programmers for the research.
- For instance, if you want to study the adoption of a new programming language, you might start with a prominent developer using that language, and they could help you identify and contact other programmers who have also adopted it.



Quota sampling

- In quota sampling, the researcher defines population classes and then determines the proportion of the population for each class.
- Quota sampling can be employed in situations where there is insufficient information about population characteristics to use probabilistic techniques. However, there is enough knowledge to create classes.





**INSTITUTO DE
COMPUTAÇÃO**



Prof. Dr. Bruno B. P. Cafeo

Sala 04
Instituto de Computação - Unicamp
Av. Albert Einstein, 1251
Cidade Universitária
Campinas – SP
13083-852

<https://ic.unicamp.br/~cafeo/>
cafeo@ic.unicamp.br