Introduction to Statistics - Young Researchers Fellowship Program

Lecture 1 - Introduction to Statistics & Tabular Data Logic

Daniel Sánchez Pazmiño

Laboratorio de Investigación para el Desarrollo del Ecuador

September 2024

What is statistics?

- A **methodology** for collecting, analyzing, interpreting, and presenting numerical information.
- A statistic is often referred to as a **numerical fact** or a **piece of data** which describes a particular characteristic of a group of individuals.
  - In the field of statistics, we typically don't refer to individual data points as statistics.
- In several fields, statistics is used as an aid to decision making under uncertainty.
- In a research context, statistics will be needed to understand phenomena, make predictions, and test hypotheses emerging from theory.
- *Statistics is the systematic investigation of the correspondence of theory with the real world*

Data in statistics

- Because statistics is concerned with information, **data** is often the starting point of any statistical analysis.
- No clear definition of data can possibly satisfy everyone, but we can think of data as a collection of **facts** to be analyzed.
  - Data is **plural** for **datum**.
- A **dataset** is a collection of data points, which can be organized in a **table**, often about a specific topic, purpose, experiment, study, or context.

Broad types of data

- Typically, statistics makes a distinction between two broad types of data:

1. **Quantitative** data, which is numerical in nature, meaning it can be measured and expressed in numbers.
   - Discrete data: whole numbers (e.g., number of students in a class).
   - Continuous data: real numbers (e.g., height, weight).
2. **Categorical** or "qualitative" data, which is non-numerical in nature, meaning it cannot be directly measured or expressed in numbers.
   - Nominal data: categories without order (e.g., colors).
   - Ordinal data: categories with order (e.g., levels of satisfaction).

How a dataset might look like