Mon 29 + Tue 30 Sept 2025

> Optional lectures on 16.30 - 18.30 SF203 from October 7

👉 What is statistics and where can we use it

→ art and science of learning from data

1. understand the data
2. get the data
3. describe the world around us

We can find 4 main components

1. RESEARCH QUESTION (RQ) = what we are interested in

   → econ, sociology, sciencepo... = something that could arise from any studying

   → now you focus on how answer and explain your RQ

2. COLLECT DATA = 2 types based on how you gather your data

   → primary = data that you gather

   ex questionnaire, survey...

   → secondary = national/international repositories where you can find proven data

   ex go to opendata.com and you get the info

1. ANALYSE DATA = understand them, summarising them, simplify them and use them
2. INTERPRET DATA = inference

What is a variable

- named whatever (x, w, y.....)
- could be whatever (characteristic, property of someone or something...)
- it varies across all the observation

2 types

1. categorical variable = each observation belong to a set (ex classes : lower, upper, middle...)

   → Nominal variables = identify one or more categories, NOT RANKABLES, ie gender

   → Ordinable variables = identify one of more categories, RANKABLES, ie how ofter do you...

2. Quantitative variable = just a number, pure data

   → Cardinal/ numerical variables = i can identify and MEASURE, ie age

Units of variables

- infinite, gathering together more than one individual
- object of the research
- individuals, age, marital status, families...
- often we give units a unique id (ie two people with the same date, but they have a number id)

2 types

1. DEPENDENT variable = it indicates the variable that the researcher wants to explain is the “effect” or the outcome = object of the research. Usually called “Y”
2. INDEPENDENT variable = it indicates the variable the researcher think its causing Y. Usually called “X”

Two types of experiment

- non experimental = you just observe → collect data

- experimental = you do something to change a situation → find new data

Ex. Emotional support in drop-out students

- non-experimental = you observe the drop-out and you see if emotional support has something to do = find data and write the study
- experimental = you build a emotional support office and you see if it has a positive benefit on drop-outs

Samples and population

→ population = EVERY single habitant, ie all 60 millions in Italy

So this is where SAMPLES come in

how do we sample? randomly, or thanks to some precedures

Quantitative info =

	Parameter	Statistics
Measure	Population notation	Sample Notation
Mean	sigma	?
Standard deviation	mi	?
	?	?

Complementary ways of analysing

To sintetize data:

1. TABLE

Identify rows and columns

→ rows = all possible categories/values for a variable (ie male/female or never/ofter/sometimes/always...)

→ columns = they identify different frequencies. different types

a. absolute frequency (AF or F) = the count of observations

b. relative frequency (RF, the weight of) or proportions (P = frequency / number of observstions =) or percentages (P% =

c. Cumulative frequency (CF) = total of one frequency (absolute, relative, percentages...). it shows the distribution at or below each categories of the value/ is the total of one (absolute or relative) frequency and all frequencies so far in the table. TOT CF is always the same as F. divided in CF (F) and CF(P%). CF CANNOT be completed in a nominal variable

→ identify all categories (IMP. with ordinal/quantitative variables you MUST respect the order)

→ identify all categories (IMP. with ordinal/quantitative variables you MUST respect the order)

quantitative variables can then be turned into classes in the table (see table on slides on bb), you decide the class but they need to be mutually exclusive

→ data can be rounded = (read slides)

2. GRAPHS