

DECEMBER 3, 2025

EVA RICCOMAGNO

ABOUT SAMPLES AND
SURVEYS

- ▶ **Setting up the topic**
- ▶ **What did they mean? Did they knew it?**
- ▶ **Public data sets**

- ▶ **Setting up the topic**
- ▶ **What did they mean? Did they knew it?**
- ▶ **Public data sets**

SURVEY vs QUESTIONNAIRE

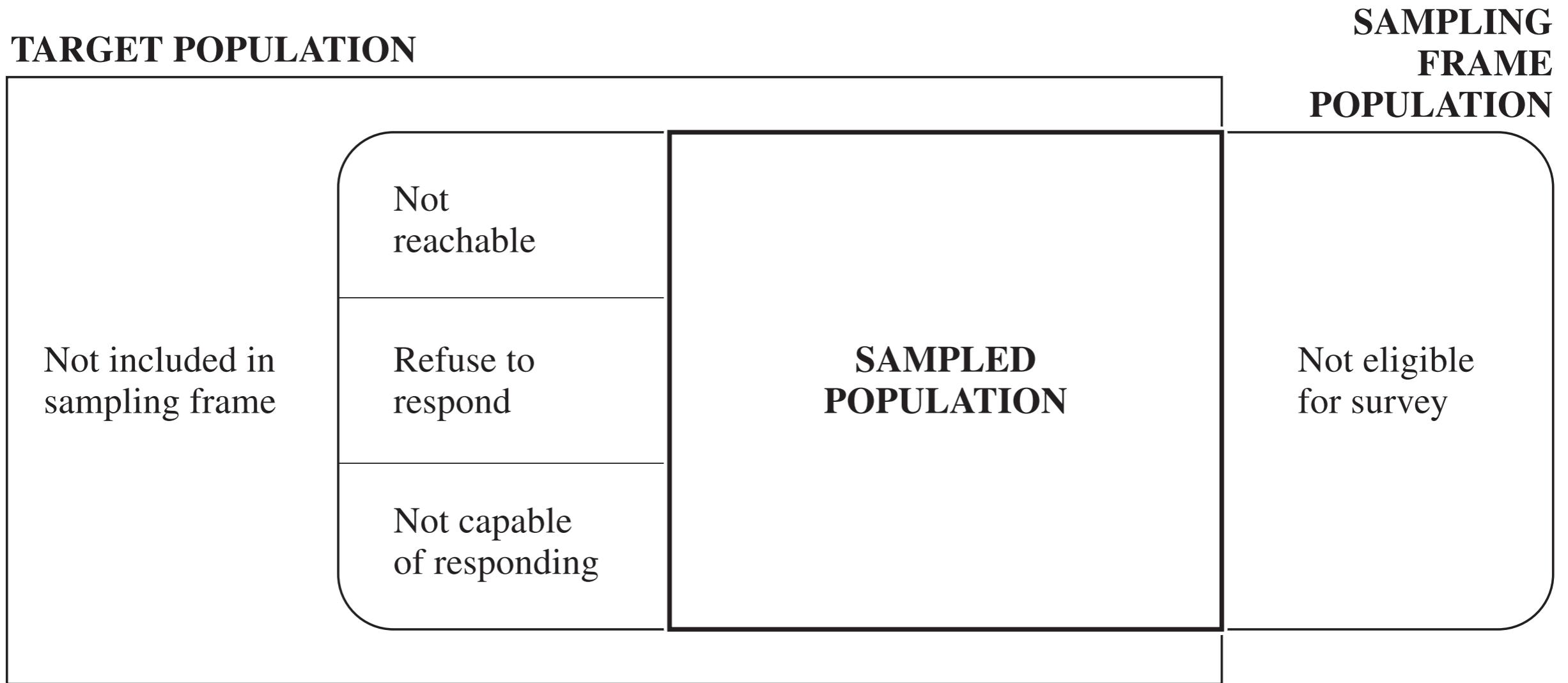
- ▶ A **survey** is a **research method** of gathering information using relevant questions from a sample of people with the aim of understanding key fact(s) of populations as a whole
 - ▶ **why WHO WHAT HOW**
- ▶ A **questionnaire** is a **specific tool or instrument** for collecting data
 - ▶ Another tools are interviews, found data, web scraping, administrative data ...

COMPONENTS OF A SURVEY

1. Survey Objective
2. Target Population
3. Sampling scheme
4. Data Collection Instrument
5. Mode of Administration
6. Data Collection Plan
7. Data Analysis
8. Reporting the Results
9. Ethical Considerations

-
- ▶ The design of a survey and its questionnaire is by far the most important aspect of survey research: no amount of statistical analysis can compensate for a badly designed survey
 - ▶ Set clear objectives
 - ▶ Determine who will participate in the survey
 - ▶ Decide the type of survey (mail, online, or in-person)
 - ▶ Design the survey questions and layout
 - ▶ Distribute the survey
 - ▶ Analyse the responses
 - ▶ Write up the results

THE WHO: population and sample



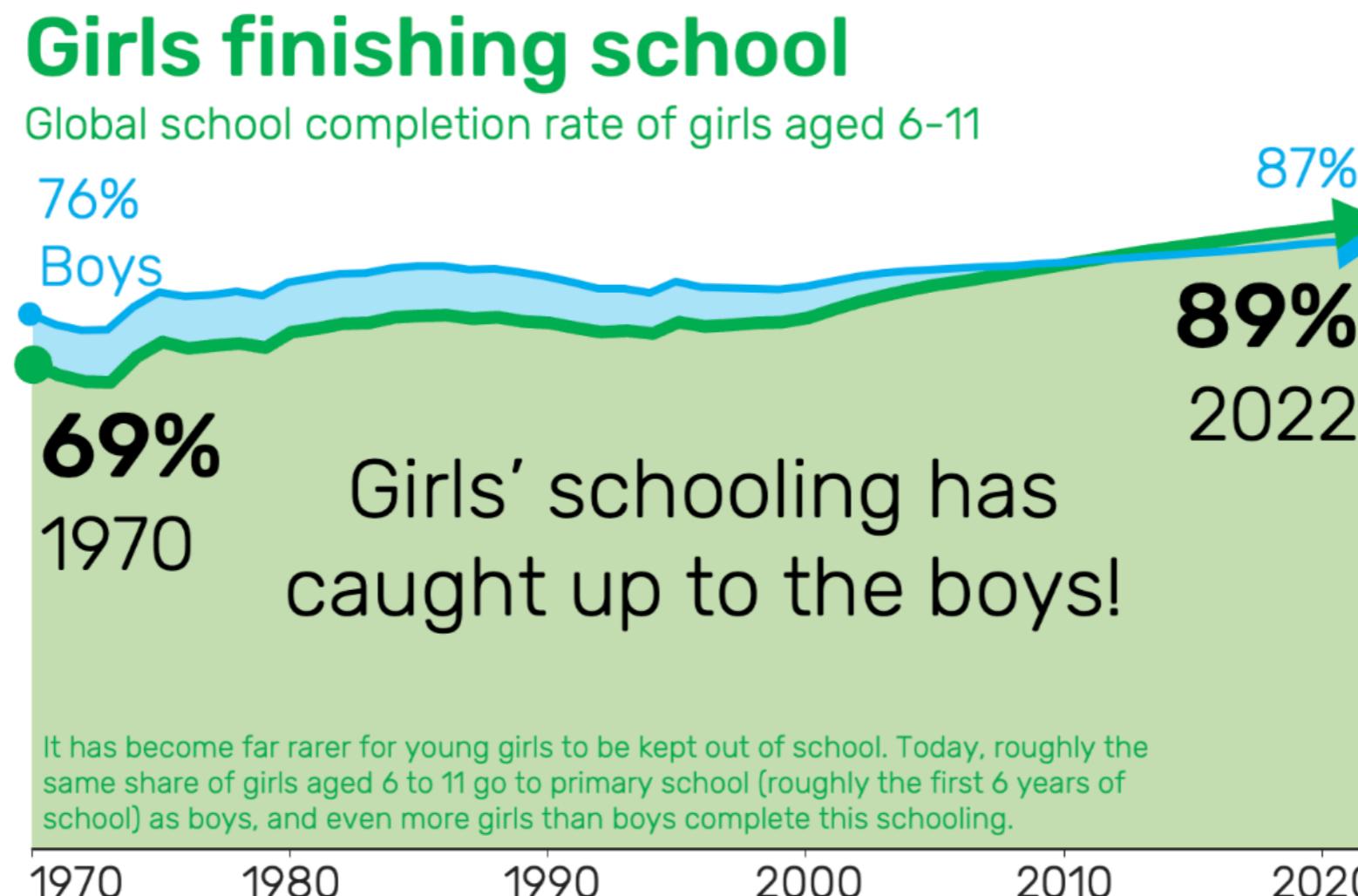
- ▶ From Sharon L. Lohr, *Sampling: design and analysis*, Second edition, 2010, Books-Cole
- ▶ Of all girls aged 6 to 11 in the world, how many go to school?

-
- ▶ Of all girls aged 6 to 11 in the world, how many go to school?
 - ▶ Less than 30%
 - ▶ Around 60%
 - ▶ Around 90%

<https://upgrader.gapminder.org/t/sdg-world-04/1755>

- Of all girls aged 6 to 11 in the world, how many go to school?
Around 90%

Worldwide, roughly nine out of ten girls go to primary school today.



Source: UNESCO

gapminder: upgrade your worldview

<https://www.gapminder.org/>

The question

Of all girls aged 6 to 11 in the world, how many go to school?

Answer options

- A: Less than 30% (**Very wrong**)
- B: Around 60% (**Wrong**)
- C: Around 90% (**Correct**)

Survey Results

Of the people we have tested, **76%** got this question wrong.

United States

24%

52%

24%

-
- ▶ A good sample should be representative in the sense that **characteristics of interest** in the population can be **estimated** from the sample with a **known degree of accuracy**

THE HOW: questionnaire design and administration

- ▶ Questionnaire can be administered electronically (by e-mail, **online**) or **in person/snail-mail**
 - ▶ large sample
 - ▶ cost in time and money
 - ▶ self-selected sample, control who's in the sample
 - ▶ anonymity
 - ▶ easy processing of data
 - ▶ time and place of data collection
 - ▶ ...
- ▶ Surroundings

THE HOW: questionnaire design and administration

- ▶ **Goal**
- ▶ **Pre-test**
- ▶ **Sentence wordings**
- ▶ **Question ordering and questionnaire layout**
- ▶ **Type of question**

THE HOW: questionnaire design - the goal

- ▶ The most important step in writing a questionnaire is deciding what you want to find out
- ▶ “I want to find out about the habits of young Europeans” won’t do
- ▶ Specific questions, such as “What percentage of 14-17 years old in Europe eat breakfast every day?” will do

THE HOW: questionnaire design - pre-test

- ▶ Always test your questions to catch misinterpretations
- ▶ Ideally on a small sample of members of the target population
- ▶ Try different versions for the questions
- ▶ Ask respondents how they interpret the questions

THE HOW: questionnaire design - pre-test (1981 Belson)

- ▶ “Do you think that the television news programmes are impartial about politics?”
- ▶ tested on 53 people

THE HOW: questionnaire design - pre-test (1981 Belson)

- ▶ “Do you think that the television news programmes are impartial about politics?”
- ▶ tested on 53 people
 - ▶ 13 interpreted the question as intended
 - ▶ 18 narrowed the term news programmes to mean “news bulletins”
 - ▶ 21 narrowed it to “political programmes”
 - ▶ 1 interpreted it as “newspapers”

THE HOW: questionnaire design - pre-test (1981 Belson)

- ▶ “Do you think that the television news programmes are **impartial** about politics?”
- ▶ tested on 53 people
 - ▶ 25 interpreted “impartial” as intended
 - ▶ 5 inferred the opposite meaning, “partial”
 - ▶ 11 as “giving too much or too little attention to”
 - ▶ the others were simply unfamiliar with the word

THE HOW: questionnaire design - sentence wordings

- ▶ Keep it simple and clear
- ▶ Avoid double negatives*
- ▶ Use specific questions instead of general ones, if possible
- ▶ Ask only one concept per question
- ▶ Avoid leading/loaded questions that prompt or motivate the respondent to say what you would like to hear
- ▶ Consider the social desirability of responses to questions, and write questions that elicit honest responses

THE HOW: questionnaire design - sentence wordings

- ▶ Many persons in the United States were shocked by the results of a 1993 Roper poll reporting that 25% of “Americans did not believe the Holocaust really happened.”
 - ▶ Question some people claim that the Nazi extermination of the Jews never happened. Have you ever heard this claim, or not?
- ▶ When the double-negative structure of the question was eliminated, and the question reworded, only 1% thought it was “possible... the Nazi extermination of the Jews never happened.”
- ▶ <https://ropercenter.cornell.edu/sites/default/files/2018-07/45031.pdf>

THE HOW: questionnaire design - sentence wordings - 2

- ▶ Do you agree or disagree... It is not true that women don't make as good executives as men do?
 - ▶ 42% 50% 7%
- ▶ Do you agree or disagree... Women make as good executives as men do?
 - ▶ 10% 87% 3%
- ▶ Burns W. Roper. “Are polls accurate?”, The Annals of the American Academy of Political and Social Science, March 1984.

THE HOW: questionnaire design - ordering and layout

- ▶ Easy, non-sensitive, closed-ended questions that will encourage the respondent to continue
- ▶ The questions should be arranged in a logical order
- ▶ If the survey covers several different topics or themes, group together related questions
 - ▶ divide a questionnaire into sections to help respondents understand what is being asked in each part
 - ▶ If a question refers back to or depends on the answer to a previous question, place it directly next to one another
- ▶ Length!

THE HOW: questionnaire design - order effect

- ▶ Serdula et al. (1995)
- ▶ Among other questions, respondents of a health survey were asked
 - ▶ to report his or her weight
 - ▶ and asked “Are you trying to lose weight?”

THE HOW: questionnaire design - order effect

- ▶ When asked to report his or her weight and then **immediately** asked “Are you trying to lose weight?”
 - ▶ 28.8% of men and 48.0% of women reported that they were trying to lose weight
- ▶ When “Are you trying to lose weight?” was asked **in the middle** of the survey and the self-report question on weight **at the end** of the survey
 - ▶ 26.5% of the men and 40.9% of the women reported that they were trying to lose weight
- ▶ The authors speculate that respondents who are reminded of their weight status may overreport trying to lose weight

THE HOW: questionnaire design - type of question

- ▶ Decide whether to use open or **closed** questions.
 - ▶ An open question allows respondents to form their own response categories
 - ▶ in a closed question (multiple choice), the respondent chooses from a set of categories read or displayed
- ▶ Each has advantages - remember: you'll have to analyse the answers
- ▶ Use forced-choice, rather than agree/disagree questions. Some persons will agree with almost any statement

THE WHAT: reporting and comparing

- ▶ Report the actual question asked
- ▶ When comparing findings from different surveys, check whether the same questions were asked (wording, order ...)
- ▶ and that the **set of possible answers** to closed questions is the same

SUMMARY

- ▶ Have clear the research question and the population of interest
- ▶ That is, what you want to find out and about whom
- ▶ Pay attention to your sample and population
- ▶ Ask clear questions with clear potential answers that are easy to analyse and informative enough to address your research question

REFERENCES

- ▶ Sharon LL, Sampling: design and analysis, Second edition, 2010, Books-Cole
- ▶ <https://www.scribbr.com/methodology/survey-research/>
- ▶ <https://www.scribbr.com/methodology/questionnaire/>
- ▶ Belson W.A, The design and understanding of survey questions, 1981, Gower Publishing
- ▶ Serdula M, Mokdad A, Pamuk E, Williamson D, and Byers T, Effects of question order on estimates of the prevalence of attempted weight loss, 1985, American Journal of Epidemiology, 142, 64–67
- ▶ Ernst, J, Democratic versus Autocratic Communications: A Comparison of American News and the German Nachricht, 1987, Communications 13, 46-54 The European Journal of Communication Research
- ▶ <https://ropercenter.cornell.edu/sites/default/files/2018-07/45031.pdf>

- ▶ **Setting up the topic**
- ▶ **What did they mean? Did they knew it?**
- ▶ **Public data sets**

... PRIMA DEL FACT CHECKING

- ▶ Lavoro di gruppo
- ▶ Scegliere la notizia
- ▶ Si capisce:
 - ▶ Quali dati
 - ▶ La popolazione di riferimento
 - ▶ Deduzioni

-
- ▶ Fonte dei dati: affidabile e riconosciuta? Dati pubblici e verificabili? Sono riportate? Accessibili?
 - ▶ Metodologia: come sono raccolti i dati? Il campione è adeguato? Sono riportati margini di errore/intervalli di confidenza?
 - ▶ Contesto e definizioni: i dati spiegano la notizia esattamente? I termini tecnici sono definiti? Il confronto è fatto tra periodi o gruppi omogenei?
 - ▶ Aggiornamento e trasparenza: o dati sono recenti? Preliminari, provvisori, definitivi?

-
- ▶ Interpretazione e possibili distorsioni: i dati sono interpretati correttamente o sono presentati in modo sensazionalistico o di parte? Correlazione e causalità sono confuse? Mancano informazioni che potrebbero cambiare la lettura del dato (e.g. cherry picking)?
 - ▶ Visualizzazione: eventuali grafici sono costruiti correttamente? Assi, scale, etichette, unità di misura...
 - ▶ Confronto con altre fonti: la notizia è riportata altrove con gli stessi dati? Ci sono smentite, verifiche, fact-checking indipendenti?
 - ▶ Ovviamente: competenza del giornalista, orientamento della testata giornalistica, conflitti di interesse, politici, commerciali o istituzionali possono colorare l'interpretazione

IN BREVE UNA NOTIZIA BASATA SU DATI E' AFFIDABILE SE

- ▶ Le fonti sono chiare e verificabili
- ▶ la metodologia è solida
- ▶ l'interpretazione è prudente e trasparente
- ▶ i grafici sono corretti
- ▶ c'è coerenza con altre fonti

CHECKLIST

- ▶ Da dove vengono i dati?
- ▶ Quale è la popolazione di riferimento? Ed il campione come si rapporta alla popolazione?
- ▶ Come sono stati raccolti i dati?
- ▶ Cosa misurano esattamente?
- ▶ I grafici/percentuali/... sono presentati correttamente?
- ▶ Altre fonti confermano?

- ▶ **Setting up the topic**
- ▶ **What did they mean? Did they knew it?**
- ▶ **Public data sets**

PUBLIC DATA SETS

- ▶ [https://www.gapminder.org/tools/#\\$chart-type=bubbles&url=v2](https://www.gapminder.org/tools/#$chart-type=bubbles&url=v2)
- ▶ **EUROSTAT**
 - ▶ <https://ec.europa.eu/eurostat/web/main/data> >data visualisation>
 - ▶ <https://ec.europa.eu/eurostat/web/education-corner/overview?language=en>
- ▶ **SISTAN (Sistema statistico nazionale)**
 - ▶ <https://www.sistan.it/>
- ▶ **ISTAT**
 - ▶ <https://www.istat.it/>
 - ▶ <https://www.istat.it/statistiche-per-temi/focus/dati-all-a-mano/>
 - ▶ <https://www.istat.it/attivita-e-servizi-per-tipo-di-utenti/studenti-e-docenti/>

► REGIONE

- ▶ <https://www.regione.liguria.it/homepage-opendata.html>
- ▶ <https://geoportal.regione.liguria.it/>
- ▶ https://lrv.regione.liguria.it/liguriass_prod/

► COMUNE

- ▶ <https://smart.comune.genova.it/opendata>
- ▶ <http://statistica.comune.genova.it/>

► ARPAL

- ▶ <https://omirl.regione.liguria.it/#/map>
- ▶ <https://www.arpal.liguria.it/tematiche/meteo/dati-osservati.html>

► Altro

- ▶ <https://www.eduscopio.it/>
- ▶