

Statistical Literacy

Problem Set 1

Rémi Viné

Due October 2nd, 2023

1. **Variables.** For each of the following variables, can you (i) determine its type, (ii) determine its level of measurement
 - (a) The height, in meters, of buildings in the city
 - (b) The level of development of countries as in table 5 of this [United Nations document](#)
 - (c) The Gross Domestic Product per capita (in USD)
 - (d) The years of education
 - (e) The highest diploma obtained (Bachelor, Master, Ph.D.)
 - (f) The SAT score
 - (g) The exchange rate (CHF / USD)
 - (h) Among the listed variables, for which ones would you be able to compute a growth rate?

2. Percentage and growth.

- (a) You are given the background education of students of the MINT program in 2023. Using table 1, find the share (the relative frequency) of each modality.

Background education	Frequency
Anthropology	43
Psychology	22
Economics	39
Political Sciences	97
Law	21
Other	48

Table 1: Frequency in 2023

- (b) Assume now that the cohort of 2022 is also available (table 2). Taking the absolute frequencies, what was the growth of the cohort per educational background? What background education has grown the most between 2022 and 2023?

Background education	Frequency in 2023	Frequency in 2022
Anthropology	43	20
Psychology	22	19
Economics	39	32
Political Sciences	97	78
Law	21	19
Other	48	51
Overall	270	219

Table 2: Frequency in 2022

3. **Descriptive, inferential.** Before a vote for a Swiss referendum on a given reform to be passed, you may want to anticipate the final outcome. You are well aware that cantons differ in their preferences and therefore, per canton, you randomly select a number of people of the biggest city in each canton. More people are selected in more populated canton.
- (a) What is the sample and what is the population of this analysis?
 - (b) What was the sampling design used?
 - (c) One conclusion of the mini-study is “About 63% of the Swiss citizens are in favor of the reform.” Is this statement a descriptive or an inferential statement?
 - (d) Can you think of possible biases affecting this analysis?
 - (e) Assume that instead of going to all Swiss cantons, you simply asked people in your surroundings. Would it change anything in the possible biases?

4. Sampling design.

- (a) Willing to understand students preferences in Europe, you wait at the cafeteria and ask every student entering the cafeteria (of IHEID) between 12.15 and 12.45 a few questions of interest. What is the population of interest? Is this procedure a probability sampling?
- (b) Now, assume that you are willing to reach conclusion to the students from the Graduate Institute only. Does it change whether or not this sampling is probabilistic?
- (c) Propose a random sampling design to study students from the Graduate Institute. In such a method, what potential biases might you expect?
- (d) Assume the direction of studies sent a questionnaire that all students have to fill up (otherwise they cannot complete their studies). It finds that “5% of all students consume illegal drugs at least once per month”. What is this “5%”, a parameter or a statistic? In this case, can you imagine some biases emerging from this?