

Final Assignment
Survey Methodology I
Due date: **January 21, 2022**

Note: You will need to wait until the last class to answer some of these questions.

1. A political analyst reviews two studies conducted before the presidential election (two candidates, A and B). The first study, nationally representative of the population registered in the electoral registers and probabilistic in all its stages, yielded a point estimate of 54.5% (valid %) in favor of candidate A. Another study using sex and age quotas and a nationwide sample of registered voters obtained a 52.3% (valid %) point estimate for candidate A. The election result was 51.6% (valid votes). The analyst notices that the study conducted by quotas was closer to the final result, raising the question of the advantages of conducting a probabilistic study (usually always more expensive).
 - a. What are the advantages and disadvantages of each type of sample? What is gained with one or the other?
 - b. What is background information needed to know the precision of the estimates (MOE) of both studies?
 - c. Why was the quota study closer to the final result?
2. A prestigious researcher was asked to study workers' opinions regarding specific measures/decisions management wanted to implement. Management hoped decisions would be well-received and wanted to anticipate the results of a referendum planned at the end of the month. The researcher's sample design was:

Group	People	%	Sample
A	2231	16,7	250
B	6450	48,3	250
C	1229	9,2	250
D	3440	25,8	250
Total	13350	100	1000

The researcher designed a survey and ran a frequency of agreement for the most controversial decision. 62% of the workers stated that they agreed with it. The researcher was sure of a favorable result, so the management conducted the referendum: the participation was high, and 96% of the workers registered their opinion. However, the final result of the plebiscite was that only 44% agreed with the measure.

- a. What type of sampling does the researcher's design correspond to?
- b. What could explain the difference in the results of the prestigious researcher, given that it is a sample of 1000 cases?

- c. Based on the answer to the previous question, is it necessary to make any adjustments to the database before running the variable frequencies? If it is necessary to make the adjustment, define how it should be calculated.
3. Suppose an important politician asks you for advice regarding the results of a poll published in a regional newspaper. According to the poll, the prominent politician wins by far (more than 20 percentage points over his primary opponent). However, based on the atmosphere on the streets, the younger and more active opponent has been adding adherents. The technical details of the study are as follows:
 - A telephone survey of **600 cases**
 - **Maximum error:** ± 4 percentage points
 - **Response rate with respect to the number of calls:** 20%
 - **Application date:** April 1-7, 2022.
 - **Strategy:** a random sample of telephone numbers of the three main providences of a region (out of 21 provinces).
 - a. What is the study's target population, and what specifications or recommendations would you give the leading politician in this regard?
 - b. Assuming that the error calculation is correct (mathematically speaking), is it appropriate to state that the maximum sampling error is ± 4 percentage points? Why? Is there a lack of background information to answer this question? If yes, what is missing?
 - c. What systematic bias would it be possible to identify, given the characteristics and performance of the study?
 - d. Finally, should the prominent politician rely on the survey results or not?
4. You were provided a dataset from a school survey to estimate adolescent drug use (column `drug` in the file `final-assignment.csv` available in the homework folder on the Github repository). No sampling weights were provided, only the marginal distribution of key variables such as gender and age group.

Marginal distribution of population

Gender: Female (0.52), Male (0.48)

Age: 9-12 years (0.40), 13-15 years (0.32), 16-18 years (0.28)

Using the techniques reviewed in the course (write your answers and attach your code or computations):

- a. Estimate the proportion of drug use
- b. Estimate the DEFF of any survey weights you compute

5. The following is a list of $A = 10$ blocks. Draw a PPS systematic sample, using X_a as the measure of size. Use a random start of 6 and an interval of 41.

Block	X_a
1	32
2	18
3	48
4	15
5	37
6	26
7	12
8	45
9	46
10	21