S 11

Data Analytics for Data Scientists

Design of Experiments (DoE)

Suggested solutions for Exercise 11: Experiments in Social Media

2025

Prof. Dr. Jürg Schwarz
MSc Adrian Bieri

Suggested solution 01

Experiments in social media – your example

Assume you are given this research question:
In a month's time, the Swiss will elect a new National Council (parliament).
The 200 party seats of the parliament are currently distributed as follows:
SVP 25.6% / Greens 21.0% / SP 16.8% / FDP 15.1% / CVP 11.4% / BDP 2.4% / Other 7.6%

What proportions can be expected in the election of the new National Council?
The research question is to be answered with data obtained from social media.

Work through the following steps – **ideally in your group** – to design a study to answer the research question:

1. Choose the design
2. Define the population
3. Reflect on population bias
4. Determine the procedure for sampling / data collection
5. Carry out the study
6. What else do you need to consider?

Create a presentation for a professional audience and explain your approach and thoughts.
The main goal is to start a discussion.
PS: You do not actually have to give a presentation. The reason for making the exercise solution a presentation is to help you learn to present your ideas in a concise manner.

Suggested answers to the questions

Note: The task has been designed very open and includes by purpose many difficulties as encountered in real researches life in various combinations.
Therefore, there are fewer "right" solutions for this task than usual.
The following section sheds light on several problematic aspects.

1 Choose a design

I choose a non-experimental observational design based on social media listening and content analysis.

- I will monitor political engagement (likes, shares, comments, hashtags, retweets) related to Swiss political parties on major social media platforms.
- I will analyse user-generated content rather than manipulating variables.

Justification:

Due to ethical and practical constraints, we cannot control or randomly assign users to political preferences; thus, we observe naturally occurring behaviour.

2 Define the population

Target population: Eligible Swiss voters (citizens aged 18+)

Accessible (social media) population:

Social media users (primarily Facebook, Instagram, TikTok, X) based in Switzerland who engage with political content related to the upcoming election.

3 Reflection on population bias:

Potential biases include:

- Coverage Bias:
  Not all eligible voters use social media or political hashtags. Older voters and rural residents might be underrepresented.
- Self-Selection Bias:
  Politically active or polarized individuals are more likely to express opinions online.
- Platform Bias:
  Younger users (e.g., TikTok) vs. older users (e.g., Facebook) might show different patterns.

Strategies for minimizing bias:

- Weight social media data by known demographic distributions (age, gender, region).
- Compare social media findings with conventional survey data.

4 Determine the procedure for sampling / data collection:

Platforms: Facebook, Instagram, TikTok, X

Procedure:

- Use social media scraping tools and official APIs where permitted.
- Search for posts containing political party names, party leader names, campaign hashtags.
- Focus on public posts and interactions only (ethical consideration).

Sampling window:

- One month prior to the election.
- Stratified by week to observe potential trends or campaign effects.

5 Implement the study:

Set up automated collection of posts and engagement metrics (likes, shares, comments retweets) using predefined search terms.

Classify posts into party categories using natural language processing (NLP) and manual checks.

Calculate relative proportions of support or mentions for each party.

Adjust results based on known demographic biases (using statistical weighting if possible).

6 What else do you need to consider?

Ethical issues:

- Respect data privacy (only public posts, anonymization of user data).

Data validity:

- Differentiate between genuine political support and negative mentions (sentiment analysis needed).

Bots and fake accounts:

- Apply bot detection algorithms to filter automated, non-human activity.

Timing effects:

- Last-minute campaign events could distort final impressions; important to monitor until election day.

External validation:

- Compare social media-based predictions to traditional opinion polls for triangulation.