

Lesson plan for participants

Summer Institute in Computational Social Science 2021

Activity, Day 4: Online Surveys in Practice

Monika Leszczynska based on activity designed by Matthew Salganik & Robin Lee [1]

Please read the entire lesson plan before beginning.

Summary

How to conduct online surveys? In this activity, you will conduct a survey using [Prolific](#) platform, analyze your data, and compare your results to those that come either from large quota samples or from a high-quality probability-based sample. We will later discuss where these differences are coming from, how to deal with them, and how important it is for legal questions to rely on probability-based samples or apply weighting to non-probability samples.

Depending on your interests, you can base your survey on one of the following:

1. Lucia A. Reisch and Cass R. Sunstein, [Do Europeans like nudges?](#)
2. [Eurobarometer](#) 91.2(2019) conducted by [GESIS – Leibniz Institute for the Social Science](#) on Europeans attitudes and knowledge about GDPR and the Charter of Fundamental Rights (Question module QB). For demographic questions, see module D of the survey.

You can also check Option 2 below for further suggestions.

Learning objectives

Participants will gain experience with the following activities:

- reading survey results and methodology reports
- creating questionnaires on Google Forms [Note: This assignment uses Google Forms because it is free. See the end of this lesson plan for comments about using [Qualtrics](#), a more sophisticated platform that requires payment or institutional support.]
- deploying surveys on Prolific
- data wrangling and, optionally, weighing non-probability samples

- using the total survey error framework to reason about and discuss sources of errors in estimates

Before group activity

- Read [Chapter 3 in Bit by Bit](#)
- Check three guides on how to conduct a study on Prolific:
 - [Creating a study](#)
 - [Google Forms Integration Guide for Prolific](#)
 - [Using attention checks as a measure of data quality](#)
- Optionally:
 - Read the paper that motivated the optional activity: [Online, Opt-in Surveys: Fast and Cheap, but are they Accurate?](#) by Goel et al. In this paper, the researchers conduct an opt-in, non-probability survey of people on MTurk. They compared the raw sample estimates and those made after various weighting adjustments to estimates that come from surveys using probability-based sampling methods.
 - Read [Intro to Poststratification](#)

Activity schedule

- 3 hours:
 - 1 hour: prepare and deploy surveys
 - 30 minutes break: data is collected on Prolific
 - 1 hour and 30 minutes:
 - 15 minutes: validate survey responses and pay Prolific participants
 - 60 minutes: analyze the dataset
 - 15 minutes: discuss activity

During group activity

1. Create a questionnaire on Google Forms. You will receive a link to a template questionnaire before the task (30 minutes).
2. Pilot test the questionnaire and check how long it will take to complete it.
3. Make sure to include demographic questions that you will need later for post-stratification (see Option 1 at the end).
4. Don't forget to include attention checks. You can follow the [guidelines](#) by Prolific.
5. Deploy your survey to Prolific (30 minutes). You can follow the guidelines described [here](#). Remember to include the completion URL at the end of the survey!
 - When setting payments for participants, please make sure that you follow Prolific's recommendations regarding fair rewards.

- We have created Prolific accounts for you to use. The accounts will be pre-loaded with funds. Please create your tasks from these accounts. If you use your own account for data collection, we won't be able to reimburse you. On Day 4, you will receive further details via the Slack channel how to access these accounts.
 - Remember to apply custom prescreening to select participants only from those countries that are of interest to your study. I would suggest to focus on the UK (the largest Prolific pool) and maybe select one more country.
6. After the data has been collected, validate the survey and pay Prolific participants (15 minutes).
- Download the CSV of responses from Google Forms.
 - Check to make sure that all of Prolific participants actually took the survey by comparing the list of Prolific IDs provided in the survey data against the Prolific IDs recorded by the Prolific platform. [2]
 - Delete responses that did not meet the attention check criteria.
 - Pay Prolific participants that completed valid surveys. When deciding to reject a submission, please follow Prolific guidelines described [here](#).
 - Remove redundant entries, if any.
 - After you have used the Prolific ID data to validate answers and to remove redundant entries, delete it from your dataset. The Prolific ID is a unique string that can be used to personally identify people.
7. Analyze the data that you collected (60 minutes).
- Compare the raw (unweighted) estimates to the results found in previous surveys. Generate plots displaying your results.
 - Discuss the differences in the results and the potential sources of these differences.
 - If you decide to run a survey based on the Eurobarometer survey, you can also do the optional tasks described below (Option 1).

Option 1

Optionally, if time allows, you can extend this exercise and produce weighted estimates, and again compare your estimates to those that come from high-quality probability-based sample. This part of the activity was inspired by a paper written by Sharad Goel, Adam Obeng, and David Rothschild: [Online, Opt-in Surveys: Fast and Cheap, but are they Accurate?](#) In this case it is recommended that you run your survey based on the Eurobarometer since the Reisch and Sunstein study did not use probability based samples.

- Download the data set from the GESIS website. You will need to login first.
- Generate the estimates from the GESIS data set.

- Next compare the estimates from the GESIS data set and from your data set after doing [cell-based post-stratification](#).
- The code [here](#) can be helpful - but please note that it was prepared for a different data set. You will need to adjust it to the data sets you are using. You will also need to get the census data for the country you decided to focus on.
- Due to the time constraint, you won't be able to do techniques as complicated in the paper by Goel et al. However, those sections are marked as optional extension in the template code, and we provide the instruction on how to do them in the code [here](#).

Option 2

You can also decide to conduct a completely different survey research than those conducted by Reisch and Sunstein or GESIS survey on the knowledge of GDPR and the Charter of Fundamental Rights. You can select the topic yourself. If you decide for this option, please follow all the steps as in the main task.

Remember that you will not be able to compare your results to those from a high-quality probability sample or a quota-based sample. Still, if you find time, you can try to conduct a cell-based post-stratification and see how your estimates change.

Additional resources

- [Here](#) are some more resources on web surveys and [here](#) you will find many additional helpful resources on how to conduct studies on Prolific.
- Dutwin and Buskirk. "[Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples](#)"
- Baker et al. "[Summary Report of the AAPOR Task Force on Non-probability Sampling](#)"
- Kalton and Flores-Cervantes "[Weighting methods](#)"
- Berinsky, Margolis, and Sances "[Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys](#)"

Appendix: Qualtrics-specific instructions

- If you are using Qualtrics instead of Google Forms, you do not have to include a question on the survey asking for Prolific ID.
- Here are [detailed instructions](#) for generating survey completion codes.

- Qualtrics has more features and customizable options -- for example, you can randomize the response options to a survey question or enable more sophisticated skip logic. Try playing around with these!

Footnotes

[1] This activity was designed with input from participants and TAs from SICSS 2017 - 2019, especially Yo-Yo Chen, Janet Xu, and Cambria Naslund.

[2] Tip: to validate Prolific ID matches, you can download a CSV of Prolific IDs from your Prolific results page and match that with your survey results data.