

Lesson plan for participants

Summer Institute in Computational Social Science AMU/Law 2022

Activity, Day 4: Online Survey

Lukasz Szoszkiewicz based on activity designed by Monika Leszczynska [SICSS-Law 2021, Maastricht University], 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 the [Prolific](#) platform and Google Forms, 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. Please note that for time efficiency we will deploy the survey on the United Kingdom population.

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

1. [Eurobarometer](#) 91.2(2019) conducted by GESIS – Leibniz Institute for the Social Science on European attitudes and knowledge about GDPR and the Charter of Fundamental Rights (Question module QB). For demographic questions, see module D of the survey. [If you choose this survey you can compare your results with the historical data]
2. [Digital rights and principles](#) (Special Eurobarometer 518) conducted by GESIS on the perception among EU citizens of the role that digital tools and the internet will play in their lives. [If you choose this survey you will only be able to compare the results against other EU countries. Due to the Brexit, the United Kingdom population was not covered in this survey].

You can also modify questions, formulate your own or merge questions from various surveys such as the [World Justice Project](#) (measuring the rule of law) or other EU-wide surveys ([see the list of topics](#)). See **Option 2** below.

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. In your future work you can use more sophisticated software such as SurveyMonkey that requires payment.]
- deploying surveys on Prolific
- data wrangling and, optionally, weighing non-probability samples

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 (suggested in the final part of this lesson plan - Option 1): “[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 the paper “[The online survey as a qualitative research tool](#)” by Braun et al. (2020). In this paper, you can find an overview and practical design information, illustrated with examples from some of the previous studies on how to ask qualitative (open-ended) questions in the online environment.

Activity schedule

- Overall: 3 hours:
 - 1 and 15 minutes: prepare and deploy the survey
 - 45 minutes break: data is collected on Prolific
 - 1 hour and 15 minutes:
 - 15 minutes: validate survey responses and pay Prolific participants
 - 60 minutes: analyze the dataset
 - 30 minutes: discuss the activity with other participants during the plenary session

During group activity

1. Create a questionnaire on Google Forms. [Here you can access a template questionnaire](#) - please copy the template on your Gmail account before implementing any changes (45 minutes).

2. Pilot the test questionnaire and check how long it will take to complete it (it should not be longer than 5-6 minutes).
3. Don't forget to include attention checks. You can follow the [guidelines](#) by Prolific.
4. 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 on how to access the account.
 - Remember to apply custom prescreening to select participants only from those countries that are of interest to your study. We suggest **focusing on the UK (the largest Prolific pool)**.
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 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 the 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 a 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 log in first.
- Generate the estimates from the GESIS data set.
- *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. The code enables to apply weighting for sex, age, region, education etc. (see line 45 for the features included). Applying weighting is necessary if you would like to generalize your findings to the whole population.
- Due to the time constraint, you won't be able to do techniques as complicated as in the paper by Goel et al. However, those sections are marked as an 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 completely different survey research than those conducted by Reisch and Sunstein or the GESIS survey on the knowledge of GDPR and the Charter of Fundamental Rights. You can select the topic yourself. If you decide on 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.

Footnotes

[1] This activity was designed with input from participants and TAs from SICSS 2017 - 2020, 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.