Sampling for web surveys

Dr. Ana Slavec, InnoRenew CoE

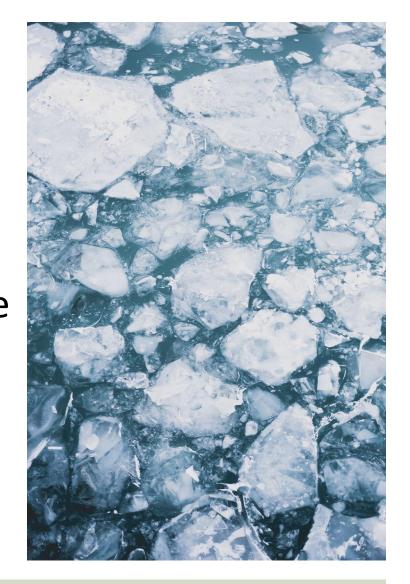ana.slavec@innorenew.eu

Statistical seminar at FAMNIT   Koper, 16. 11. 2021

Icebreaker questions

- Have you participated in a survey in the last year/month?

- How were you invited to participate in the survey?

- Have you ever created a survey?

- How did you invite participants?

Outline of lecture

- I. Surveys as a social science method

- II. Survey sampling

- III. Sampling for web surveys

- IV. Other considerations for web surveys
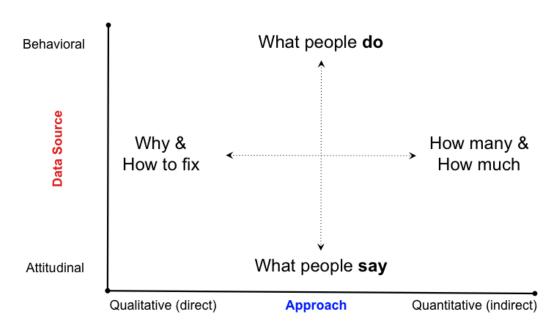
I. Surveys as a social science method

- Quantitative research that uses survey methodology to answer questions about a population

- Used for public-opinion polls, market research, health-care research, psychology, sociology, etc.

Qualitative vs quantitative research

- Complex
- Intensive
- Focused on cases/units
- A few cases

Questions answered by research methods based on **Data Source** & **Approach**

Behavioral

Data Source

What people **do**

Why & How to fix

How many & How much

Attitudinal

What people **say**

Qualitative (direct)   **Approach**   Quantitative (indirect)

© 2008 by Christian Rohrer

- General
- Extensive
- Focused on attributes/ dimensions
- A lot of cases

Designed vs. Organic Data

Designed data (active)

- Experiments, Surveys

Organic data (passive)

- Administrative, Transactional

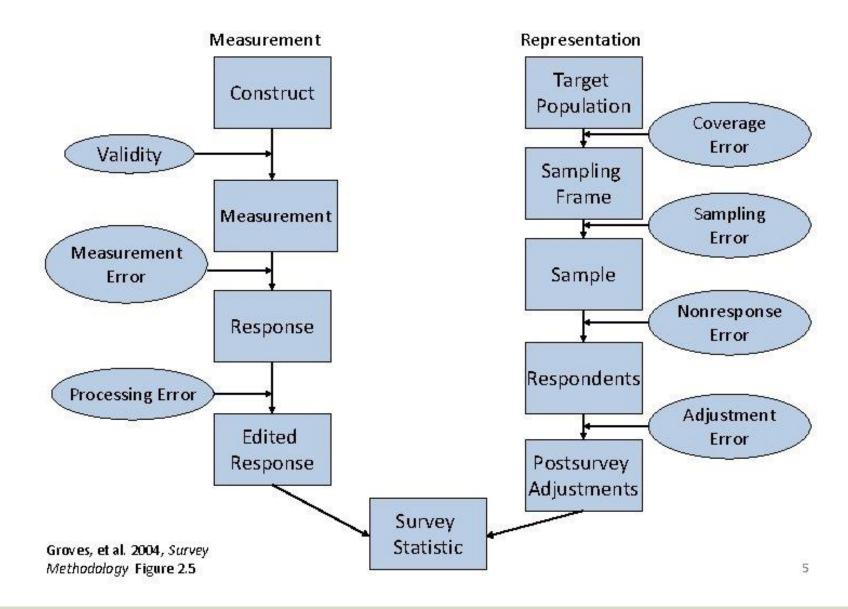# Three eras of survey research (Groves 2011)

```
1930s
Invention of sampling statistics
```
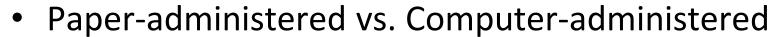
```
1960s
Expansion of use with advances
in technology (computer
assisted telephone surveys)
```

```
1990s Supplementation with
organic data
```

Survey process

Groves, et al. 2004, *Survey Methodology* Figure 2.5

Survey mode

- Interviewer-administered vs. Self-administered
  - Face-to-face survey 🤝
  - Telephone surveys 📞
  - Mail surveys 📭
  - Web surveys 🌐

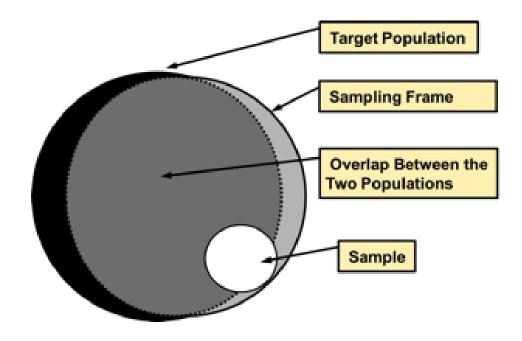- Paper-administered vs. Computer-administered
  - Paper and pencil ✏️
  - Computer administered 💻
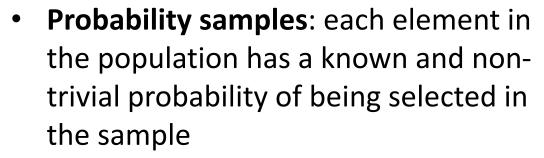
II. Survey sampling

- **Sampling:** the process of selecting a part of the population to be included in the survey

- **Sampling frame**: list of elements (units) of the general population from which we select the elements in a sample

Source: www.theanalysisfactor.com

Sampling methods

- **Probability samples**: each element in the population has a known and non-trivial probability of being selected in the sample
    - Simple random sampling
    - Systematic sampling
    - Stratified sampling
    - Cluster sampling

- **Non-probability samples:** the probability of selection cannot be calculated
    - Casual samples
    - Expert choice
    - Quota sampling

Note: Statistical inference can be used only for probabilistic samples.

Sample size

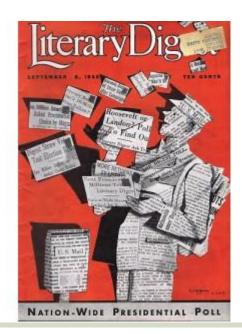|	Confidence level = 95%			Confidence level = 99%		
	Margin of error			Margin of error		
Population size	5%	2,5%	1%	5%	2,5%	1%
100	80	94	99	87	96	99
500	217	377	475	285	421	485
1.000	278	606	906	399	727	943
10.000	370	1.332	4.899	622	2.098	6.239
100.000	383	1.513	8.762	659	2.585	14.227
500.000	384	1.532	9.423	663	2.640	16.055
1.000.000	384	1.534	9.512	663	2.647	16.317

← Accuracy
← Precision

Example: US 1936 election

Literary digest magazine:

- 10 million questionnaires mailed
- 2.4 million returned
- Prediction: 57% Landon

George Gallup:

- Quota sample of 50,000 people
- Control over demographic characteristics
- Prediction: 56% Roosevelt

Franklin Roosevelt won with 62% votes
(Gallup was only 6 percentage points wrong)
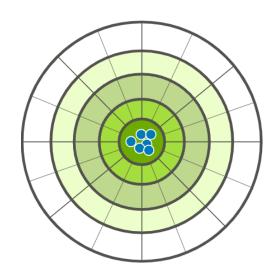
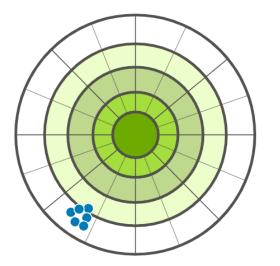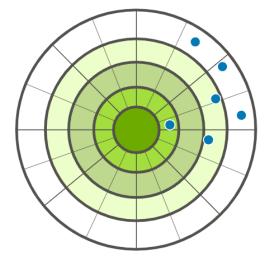**Size is not all that matters**

Source: imgur.com

Accuracy vs precision

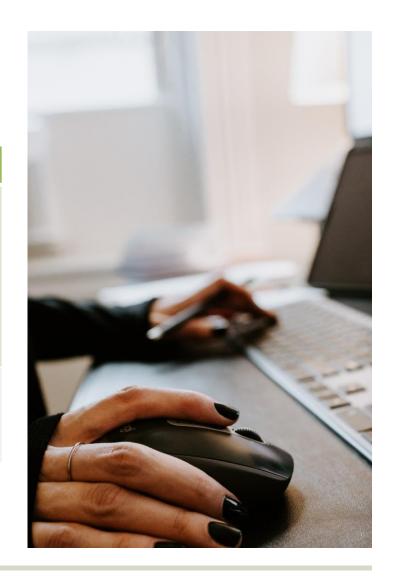Accurate AND Precise   Accurate NOT Precise   Precise NOT Accurate   NOT Precise or Accura
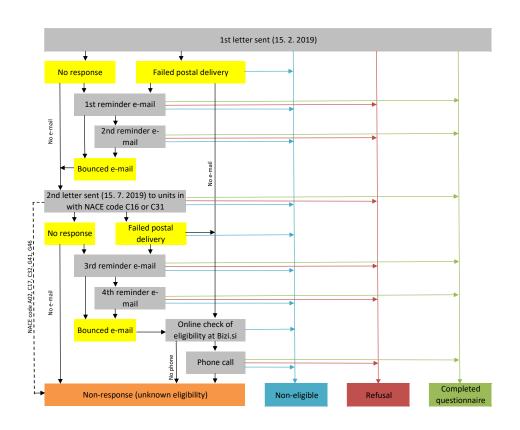
Source: publicdomainvectors.org

III. Sampling for web surveys

	Probability sampling	Non-probability sampling
List-based surveys	• Surveys of specific Internet populations • **Surveys of the general population (mixed mode)** • Probability-based online panels	• Survey with incomplete list of the target population • Survey based on list collected by self-selection • **Non-probability online panel**
Non-list based surveys	• Intercept surveys	• **Unrestricted web survey (self-selection)**

Example of mixed-mode survey with internet-based option

Survey on Innovation Activities in the Wood-Based Value Chain (Slavec and Burnard 2021)

Example of two case studies of running two parallel non-probability surveys (Unrestricted non-list based vs. List based panel)

1. Young doctors of Slovenia's **survey on compliance with COVID-19 protective measures** (November 2020) 😷

2. University of Maribor student's **survey on COVID-19 vaccination attitudes** (December 2020) 💉

Sample

	Large convenience sample (non-list based, unrestricted, snowball sample, shared on social media)	Smaller representative sample (list-based, non-probability panel from marketing research company Valicon)*
😮 Case study 1: COVID-19 protection measures	5,165 responses	527 responses
💉 Case study 2: COVID-19 vaccination attitudes)	12,042 responses**	516 responses

* It is the largest online panel in Slovenia and three quarters of panalists were included based on **probability sampling**. The representativness of the sample is controled with comparisons with offical data.

We have a list of contacted organisations and we estimated the potential reach to **more than 400,000 users (assuming a zero overlap in followers) which is about 48 % of all social media users according to the data of the Statistical office of Slovenia. However, organic posts usually do not reach all users and according to some estimates the average reach is only between 5-6% of the followers, i.e. between 20,000 and 24,000. However, since there were **45,633 unique clicks** on the survey link in the eleven days that the survey ran, the reach was probably higher. We estimated the response rate is between 3 % (assuming the survey link reached more than 400,000 people) and 26 % (assuming the reach was the same as the number of clicks).

Note: all four databases were weighted by gender, age, education and region

Case study 1: COVID-19 protection measures 🥴

	Large sample (non-list based)	Smaller sample (list-based)
% that had two or more symptoms and/or were in a risky contact in the last 7 days	10% of respondents	21% of respondents
% that have been at workplace at least once in the last 7 days	57% of employed respondents	64% of employed respondents
% that visited at least one other household in the last 7 days	44% of respondents	47% of respondents
% wore masks correctly on visits in the last 7 days	59 % of visiting respondents	44 % of visiting respondents
Differences between regions	Sig. differences for almost all indicators (e.g. % that visited other household ranging from 37% to 58%)	Regions too small to calculate differences

Case study 2: COVID-19 vaccination attitudes

	Large sample (non-list based)	Smaller sample (list-based)
% that will probably or definitely get vaccinated	59% of respondents	50% of respondents
% of medical professionals	17% of respondents	4% of respondents
Differences in vaccination intention between healthcare professionals	Significant differences in between professions (84% physicians and 82% medical students intend to get vaccinated compared to only 51% healthcare students and 50% other healtchare professions)	Not enough healthcare professionals included in sample

More about the results of the two studies

Compliance with COVID-19 protective measures 😷

- Medium posts (in Slovenian):
  - Večina še vedno na delovnih mestih
  - Obiski in širjenje virusa
  - Kako upoštevamo zaščitna priporočila?
  - Bolj verjetni prenašalci virusa
  - Samozaščitni ukrepi – razlike med regijami?

- Data in the process of being deposited to the Slovenian Social Science Data Archive

Attitudes towards COVID-19 vaccination 💉

- Medium posts (in Slovenian):
  - Koliko ljudi se namerava cepiti?
  - Koliko zdravstvenih delavcev se bo cepilo?
  - Dejavniki, ki vplivajo na odnos do cepljenja proti COVID-19

- Research paper (in English): DOI

vaccines
an Open Access Journal by MDPI
Factors Affecting Attitudes towards COVID-19 Vaccination: An Online Survey in Slovenia
Luka Petravić; Rok Arh; Tina Gabrovec; Lucija Jazbec; Nika Rupčić; Nina Starešinič; Lea Zorman; Ajda Pretnar; Andrej Srakar; Matjaž Zwitter; Ana Slavec
Vaccines 2021, Volume 9, Issue 3, 247

- Data available through the Slovenian Social Science Archive: SARSPR20, SARSVE20

IV. Other considerations

- Survey burden
- Survey introduction
- Question wording
- Questionnaire design
- Questionnaire pre-testing

Survey introduction

- Clear identification of research affiliation and funding

- Presentation of aim and relevance of research

- Clearly visible URL

- Estimate of survey length

- Ensuring confidentiality and privacy (GDPR), informed consent

- Contact information

Survey burden

- **Satisficing** is a decision-making strategy that aims for a satisfactory or adequate results, rather than the optimal solution (Krosnick 1991)

$$p = \frac{difficulty\ of\ task}{motivation * ability}$$

- Questionnaire design should strive to:
  - Lower the difficulty of task
  - Increase the motivation of the respondent

Survey question wording

Two priests are discussing if it is a sin to smoke and pray at the same time. After failing to reach a conclusion, each goes off to consult his respective superior.

The next day they meet again.

A: „What did your superior say?"

B: „He said it is all right."

A. „That's funny. My superior said it was a sin."

B: „What did you ask him?"

A: „I asked him if it was all right to smoke while praying"

B: „Oh. I asked my superior if it was all right to pray while smoking."

Source: Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires

(Bradburn, Sudman & Wansking, 2004)

Typical questionnaire design mistakes

- Too many questions per page
- Too many complex questions
- Non-application of conditions
- Non-appropriate format of questions
- Answer options not mutually exclusive
- Answer options not comprehensive
- Multidimensionality of questions

Questionnaire pre-testing methods

- Expert evaluation
- Cognitive interviews
  - Think-aloud technique
  - Paraphrasing
- Focus groups
- Field testing
  - Response quality indicators
  - Split-ballot experiments
- Computerised models
  - Survey Quality Predictor (SQP)
  - Question Understanding AID (QUAID)
  - Text corpora approach (Slavec and Vehovar 2016)

Data sharing

- Pre-registration, reproducibility
  - Center for Open Science
  - Leibniz Insitute for Psychology information
    - Michael Bosnjak: Open Science and Services for Your Scientific Work

- Data Arhives
  - CESSDA – Consortium of European Social Science Data Archives
  - GESIS – Leibniz Institute for the Social Sciences
  - ADP - Slovenian Social Science Data Archive
    - Event tomorrow: Social science research on aspects of the Covid-19 pandemic in Slovenia
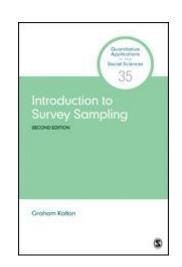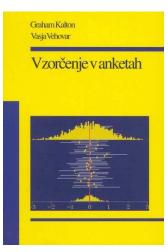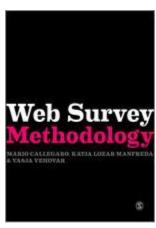
Further reading

Graham Kalton. <u>Introduction to survey sampling</u>, 2nd edition. Sage, 2020.

Graham Kalton & Vasja Vehovar. <u>Vzorčenje v anketah</u>. Založba FDV, 2011.

Mario Callegaro, Katja Lozar Manfreda & Vasja Vehovar. <u>Web Survey Methodology</u>. Sage, 2015.

Questions?

Thank you for your attention.