

Understanding Survey Data and Public Polls

Survey Research

Survey research as a research methodology (comparable to experimental research, ethnography, grounded theory, case studies, etc.) relies on questionnaires or interviews for data collection. It can employ quantitative research strategies (e.g., using questionnaires with numerically rated items), qualitative research strategies (e.g., using open-ended questions), or both (i.e., mixed methods). Surveys are crucial for gathering data in social sciences, market research, public health, and policy-making. They help researchers understand public opinion, behaviors, and trends, and inform decisions based on empirical evidence.

Key Concepts in Survey Research

Several concepts are critical to understanding survey data and assessing its quality:

- **Target Population** is the entire group of individuals or entities to which the survey results will be generalized. This group should be clearly defined based on the research objectives.
- **Sampling Frame** is the list or database from which the sample will be drawn. It should be as comprehensive and up-to-date as possible to accurately represent the target population. A perfect sampling frame is complete and includes all elements or sample units from the population, with each element listed once and only once, and without any irrelevant or extraneous elements.
- **Sample Size** refers to the number of individuals or units selected from the sample frame that is included in a specific study. A larger sample size generally increases the study's statistical power, reduces the margin of error, and provides more confidence in the findings.
- **Sampling Method** is the technique used to select a sample from a population, and it directly affects

the accuracy, reliability, and representativeness of survey results, with probability sampling methods generally providing higher quality data compared to non-probability methods.

» Probability Sampling Methods

- » **Simple Random Sampling:** Every member of the sampling frame has an equal chance of being selected.
- » **Systematic Sampling:** Individuals are selected at regular intervals from the sampling frame.
- » **Stratified Sampling:** The population is divided into subgroups (strata) based on certain characteristics (e.g., age, gender, income level, geographic region), and samples are drawn from each stratum. Stratified sampling expects that the measurement of interest varies between the different subgroups and enhances representativeness by ensuring that key subgroups are proportionately represented.
- » **Cluster Sampling:** The population is divided into clusters, and a random sample of clusters is selected. All individuals within the chosen clusters are surveyed. Cluster sampling is cost-effective and practical for large, dispersed populations, though it may introduce cluster-related bias if clusters are not homogeneous.

» Non-Probability Sampling Methods

- » **Convenience Sampling:** Samples are selected based on ease of access or proximity to the researcher. It is a quick and inexpensive choice but often results in a biased and unrepresentative sample, reducing the generalizability of the survey results.
- » **Quota Sampling:** Samples are selected to ensure certain characteristics are represented in specific proportions. For example, a market researcher decides to survey 750 people over 20 years old and set quotas to ensure that the sample resembles the proportion in the U.S. population: 270 individuals aged 20-39, 256 aged 40-59, and 224 aged 60 and above. Quota sampling

does not use random selection within each subgroup. Instead, the researcher selects readily available individuals who meet the quota criteria. The non-random selection can introduce bias.

» **Judgment (or Purposive) Sampling:**

Samples are selected based on the researcher's judgment about which units will be most useful or representative. It is a selective and subjective sampling.

- » **Snowball Sampling:** Samples are selected by asking the participants to nominate subjects known to them. This method is commonly used when investigating hard-to-reach groups. However, snowball sampling is subject to bias due to the lack of control over recruitment and those with more connections are more likely to be included.

- **Data Collection Mode** refers to the method or approach used to collect information from respondents in a survey. Common modes include online surveys, telephone surveys, face-to-face surveys, mail surveys, and mixed-mode surveys.
- **Survey Instruments** refer to specific tools or mediums used to collect information from respondents. Common survey instruments include:
 - » **Paper Mailout-Mailback Instrument:** A traditional paper questionnaire sent to respondents via postal mail, which they complete and mail back.
 - » **Self-Administered Questionnaire (SAQ):** A paper or electronic questionnaire that respondents complete on their own without an interviewer.
 - » **Face-to-Face Interview:** A structured or semi-structured questionnaire administered in person by an interviewer.
 - » **Computer-Assisted Personal Interview (CAPI):** An interview conducted face-to-face using a tablet or laptop, where the interviewer enters responses directly into the computer.
 - » **Computer-Assisted Telephone Interview (CATI):** An interview conducted over the phone with the aid of a computer system that guides the interviewer through the questionnaire and records responses.

» **Diary Methods:** Participants keep a diary or log of activities, behaviors, or experiences over a period of time.

- **Nonresponse Followup (NRFU)** is a process used in surveys to address and reduce the impact of nonresponse, where some respondents do not initially participate or complete the survey. It involves reaching out to nonrespondents through various methods, such as additional reminders, alternative contact methods, incentives, and in-person visits.

Types of Surveys

Surveys can be categorized into different types based on their design and data collection approach:

- **Cross-Sectional Survey:** Collects data from a representative sample of respondents at a single point in time, providing a snapshot of the population's characteristics or opinions at that moment.
- **Longitudinal Survey:** Tracks the same individuals or group of respondents over multiple points in time, allowing for the study of changes and developments within the sample over a period.



Survey Data Examples

Survey	Collection Approach	Target Population	Sampling Frame	Sample Size	Sampling Method	Survey Instrument	Data Products
American Community Survey	Cross-sectional survey	The U.S. population	The Census Bureau’s Master Address File (updated twice a year with the USPS Delivery Sequence File)	1,980,550 housing units in 2022; 124,846 group quarter people	Multi-stage probability sampling	Internet self-administered questionnaire; paper instrument; CAPI follow-up for a sample of nonrespondents (NRFU).	Census American Community Survey data products
IPSOS Knowledge Panel	Longitudinal panel (by invitation) with an additional online opt-in panel	Adult U.S. population	60,000 random sampled panel members drawn from the USPS Delivery Sequence File frame.	Sample size varies for each survey	Address-based probability sampling	Only online questionnaire (it provides non-internet households with a tablet and mobile data plan)	IPSOS Poll
American Trends Panel (managed by IPSOS)	Longitudinal panel (by invitation)	Adult U.S. population	12,000 adult panel members drawn from the USPS Delivery Sequence File frame	Sample size varies (a subset of panelists)	Address-based probability sampling	Since 2016 online-only panel	Pew Research Center Survey Data
National Consumer Panel (NCP)	Opt-in Panel	U.S. households	Over 120,000 households in the Panel profile (among which, 46-52% is in its Static Panel)	n/a	Non-probability convenience sampling	Scanning equipment, or NCPMobile App to transmit shopping data	Data feeds Circana (formerly IRI) and NielsenIQ
YouGov	Opt-in Panel	Total population	27+ million registered panel members worldwide	1,500+ for each poll	Non-probability convenience sampling (responses weighted to be representative of the full population)	Online questionnaires	Data feeds the New York Times; and CBS News public polls.



To learn more, refer to **Data Quality Literacy Series 13: Evaluating Survey Data Quality.**

References

Census Bureau. (2022). *American Community Survey and Puerto Rico Community Survey Design and Methodology*. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/2022/acs_design_methodology_report_2022.pdf

Ipsos. (n.d.). *KnowledgePanel: A Methodological Overview*. <https://www.ipsos.com/sites/default/files/ipsosknowledgepanelmethodology.pdf>

Johnson, R. B., & Christensen, L. (2017). Methods of Data Collection in Quantitative, Qualitative and Mixed research. *Educational Research: Quantitative, Qualitative and Mixed Approaches*, 179-206. https://uk.sagepub.com/sites/default/files/upm-assets/106363_book_item_106363.pdf

Liu, G., Bordelon, B., Nagar, R., Sarti, J., Nguyen, U., & Boettcher, J. (2024). *Data Quality Literacy: A Guidebook*. Institute of Museum and Library Services (IMLS) Grant Project. <https://doi.org/10.31219/osf.io/ruawm>

Pew Research Center. (2014). *Q&A: What the New York Times' Polling Decision Means*. <https://www.pewresearch.org/short-reads/2014/07/28/qa-what-the-new-york-times-polling-decision-means/>

Pew Research Center. (2024). *American Trend Panel*. <https://www.pewresearch.org/the-american-trends-panel/>

Ponto, J. (2015). Understanding and Evaluating Survey Research. *Journal of the Advanced Practitioner In Oncology*, 6(2), 168. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4601897/>

YouGov. (n.d.). *Panel Methodology*. <https://today.yougov.com/about/panel-methodology>

The Knowledge Brief is compiled by Grace Liu, inspired by the National Forum presentation from John M. Abowd and Kathleen Weldon, reviewed by the IMLS Data Quality Literacy project team, and designed by Niko Galimoto. This project was made possible in part by the Institute of Museum and Library Services [RE-252357-OLS-22].

Visit the project website to learn more!
<https://www.dataqualityliteracy.org>