



6.1 Introduction to Survey Design

A survey is a systematic method of collecting information from individuals to gain insights into opinions, behaviors, or characteristics. Surveys are essential in fields such as market research, social sciences, healthcare, and customer feedback analysis. A well-designed survey is crucial for ensuring the reliability, validity, and usefulness of the collected data.

Principles of Good Survey Design

1. Clear Objectives

- Define the purpose of your survey and the specific data you aim to collect.

2. Know Your Audience

- Understand the demographic and cultural context of your respondents to ensure relevance and clarity.

3. Question Clarity

- Use simple, unambiguous language.
- Avoid leading or double-barreled questions that may confuse respondents or bias their answers.

4. Survey Length

- Keep it concise to maintain engagement while ensuring comprehensive data collection.

Types of Surveys

Surveys can be categorized based on their method, purpose, or frequency:

A. By Data Collection Method

1. Online Surveys

- Distributed via email, social media, or platforms like Google Forms.
- **Advantages:** Cost-effective, scalable, and fast.

2. Paper Surveys

- Suitable for populations with limited internet access.
- **Use Case:** Rural or underconnected regions.

3. Telephone Surveys

- Effective for quick responses but prone to low response rates.

4. Face-to-Face Interviews

- Provide in-depth insights but are time-intensive and costly.

B. By Purpose

1. Descriptive Surveys

- Describe characteristics or opinions of a population.

- **Example:** Customer satisfaction surveys.

2. Analytical Surveys

- Explore relationships between variables.
- **Example:** Surveys linking exercise habits to mental health.

3. Exploratory Surveys

- Generate ideas or gather initial insights.
- **Example:** Brainstorming new product features.

C. By Frequency

1. Cross-sectional Surveys

- Conducted at a single point in time to capture a snapshot of opinions or behaviors.
- **Example:** Dietary habits survey among college students during one semester.
- **Purpose:** Understand behaviors at a specific time.

2. Longitudinal Surveys

- Repeatedly conducted over time to track changes or trends.
 - **Example:** Annual health studies tracking exercise habits.
 - **Purpose:** Capture trends and behavioral evolution.
-

Types of Questions

1. Open-ended Questions

- Allow respondents to answer in their own words, providing detailed insights.

2. Closed-ended Questions

- Include multiple-choice, yes/no, or ranking formats.

3. Rating Scales

- Commonly use Likert scales (e.g., strongly agree to strongly disagree).
-

Sampling Methods

Probability Sampling

1. Simple Random Sampling (SRS)

- **Description:** In SRS, every individual in the population has an equal chance of being selected. This is typically achieved using random number generation or drawing names from a hat.
- **Purpose:**
 - To obtain an unbiased sample when the population is homogeneous, or when there is a list of all members.
- **Example:**
 - A university wants to select 100 students from its 1,000 students for a survey on campus life. By randomly choosing 100 students from the list, every student has an equal chance of being selected, ensuring no bias in the sample.

2. Systematic Sampling

- **Description:** In systematic sampling, every n th individual is selected from a list after choosing a random starting point. The interval between selections (k) is typically determined by dividing the population size by the desired sample size.
- **Purpose:**
 - To simplify the sampling process when the population list is ordered and when random sampling is not feasible due to time or resources.
- **Example:**
 - A company wants to survey 200 employees about job satisfaction but has a list of 1,000 employees. They choose a random starting point (say, employee 5) and then select every 5th employee on the list to form the sample.

3. Stratified Sampling

- **Description:** The population is divided into mutually exclusive subgroups (strata) based on a certain characteristic (e.g., age, gender, income). A random sample is taken from each stratum.
- **Purpose:**
 - To ensure that specific subgroups of the population are represented proportionally, especially when the population is heterogeneous.
- **Example:**
 - A health survey wants to study the habits of different age groups. The population is divided into strata such as 18-30, 31-45, 46-60, and 60+. A random sample is taken from each age group to ensure all ages are represented.

4. Cluster Sampling

- **Description:** The population is divided into clusters (often based on geography or other natural groupings), and a random sample of clusters is selected. All individuals within the chosen clusters are then surveyed.
- **Purpose:**
 - To reduce costs and logistical challenges when the population is large and dispersed over a wide area, making individual selection impractical.
- **Example:**
 - A government wants to conduct a census of rural households in a country. Instead of surveying every household, they randomly select a few villages (clusters) and survey all households within those selected villages.

5. Multistage Sampling

- **Description:** Multistage sampling is a combination of various sampling techniques. For example, it may involve cluster sampling at the first stage, followed by simple random sampling or stratified sampling in subsequent stages.
- **Purpose:**
 - To handle large and geographically dispersed populations by breaking down the sampling process into manageable stages.
- **Example:**
 - A research organization conducting a nationwide health survey first selects random regions (clusters), then selects random households within those regions, and finally surveys individuals within those households.

6. Proportional vs. Equal Allocation in Stratified Sampling

- **Description:**
 - **Proportional allocation:** The sample size from each stratum is proportional to the stratum's size in the population.
 - **Equal allocation:** The same sample size is taken from each stratum, regardless of its size in the population.
- **Purpose:**
 - **Proportional allocation:** To ensure the sample reflects the population's structure, with larger groups contributing more to the sample.
 - **Equal allocation:** To treat each subgroup as equally important, even if some subgroups are smaller than others.
- **Example:**
 - **Proportional:** A company conducting a market survey about customer satisfaction with three regions (North, South, West), with the South region having 50% of customers, North 30%, and West 20%. In a proportional sample, the South region would contribute 50% of the total sample.
 - **Equal:** A researcher studying voter preferences across different political parties in a country might want to sample an equal number of people from each party, regardless of the size of the party in the population.

Key Considerations for Choosing a Sampling Method:

- **Population Size and Access:** Consider how easy it is to access and list the population. For small or accessible populations, SRS might be appropriate. For large, dispersed populations, cluster or multistage sampling might be more cost-effective.
- **Homogeneity vs. Heterogeneity:** If the population is highly heterogeneous (e.g., differing in income or education levels), stratified sampling can ensure all subgroups are properly represented.
- **Cost and Time:** Some methods (e.g., cluster sampling, multistage sampling) may be more cost-effective, especially for large populations or geographically dispersed groups.
- **Precision vs. Practicality:** Stratified and proportional sampling generally give more accurate results, but simpler methods like systematic sampling may be more practical when time or resources are limited.

Each probabilistic method has its advantages, depending on the nature of the population and the goals of the survey.

Non-probability Sampling

- Includes convenience, quota, or snowball sampling.
- **Caution:** May introduce biases due to non-random selection.

Common Survey Biases

1. Response Bias

- Respondents may give socially desirable or inaccurate answers.
- **Example:** Overstating healthy eating habits.

2. Selection Bias

- The sample does not represent the target population.
- **Example:** Internet usage survey sampling only urban residents.

3. Non-response Bias

- Certain groups are underrepresented due to low participation.
- **Example:** Workplace satisfaction survey where only satisfied employees respond.

4. Acquiescence Bias

- Tendency to agree with statements regardless of actual feelings.
- **Example:** Automatically selecting “agree” in satisfaction surveys.

Strategies to Minimize Biases:

- Conduct pilot testing.
 - Use clear instructions.
 - Offer incentives to encourage participation.
-

Tools for Survey Creation

- **Google Forms**
- **Microsoft Forms**
- **Survey Monkey**
- **Quadratics**

Features:

- Customization templates, real-time analytics, and easy sharing.
-

Ethical Considerations

1. Informed Consent

- Ensure participants understand the survey’s purpose and data usage.

2. Confidentiality

- Protect respondents’ identities and sensitive information.

3. Voluntary Participation

- Make it clear that participation is optional, and respondents may skip questions if desired.

Here’s a section that you can include in your lecture to clarify how to determine the sample size for a survey:

Determining the Sample Size for Surveys

Determining the appropriate sample size is a crucial step in designing a survey, as it directly impacts the accuracy and reliability of the results. If the sample is too small, the survey may not capture enough diversity or may lead to misleading conclusions. If the sample is too large, it can be inefficient and costly. Here’s a breakdown of the key factors to consider when determining sample size:

Key Factors in Sample Size Calculation

1. Population Size (N):

- The total number of individuals in the population you are studying. If the population is large (e.g., millions of people), it has less impact on the sample size calculation. However, for small populations, this factor becomes more significant.

2. Margin of Error (E):

- The margin of error represents the amount of random sampling error you are willing to tolerate in your results. It is typically expressed as a percentage (e.g., $\pm 5\%$). Smaller margins of error require larger sample sizes.
- Example: A margin of error of $\pm 5\%$ means the true value of the population parameter is likely within 5% of the sample estimate.

3. Confidence Level (Z):

- The confidence level reflects how confident you want to be that the sample accurately represents the population. Common confidence levels are 90%, 95%, and 99%. The higher the confidence level, the larger the required sample size.
- The Z-value corresponds to the desired confidence level (e.g., for 95% confidence, $Z = 1.96$).

4. Population Proportion (P):

- This refers to the estimated proportion of the population that is expected to exhibit a certain characteristic (e.g., 50% of the population supports a policy). If you don't know this proportion, you often assume 50% (0.5), which gives the maximum variability and ensures the largest possible sample size.

Sample Size Formula

For a simple random sample, the sample size (n) can be calculated using the following formula:

$$n = \frac{Z^2 \cdot P \cdot (1 - P)}{E^2}$$

Where: - **Z** is the Z-score associated with the chosen confidence level. - **P** is the estimated population proportion. - **E** is the desired margin of error.

Step-by-Step Example

Let's calculate the sample size for a survey with the following parameters: - **Population size (N)** = 10,000 - **Confidence level** = 95% ($Z = 1.96$) - **Estimated population proportion (P)** = 50% or 0.5 (most conservative estimate) - **Margin of error (E)** = $\pm 5\%$ or 0.05

Using the formula:

$$n = \frac{(1.96)^2 \cdot 0.5 \cdot (1 - 0.5)}{(0.05)^2}$$

$$n = \frac{3.8416 \cdot 0.25}{0.0025}$$

$$n = 384.16$$

So, the sample size needed for a 95% confidence level with a 5% margin of error would be **384** participants.

Adjusting for Small Populations (Finite Population Correction)

If your population is relatively small (say less than 20,000), you can apply a finite population correction (FPC) to adjust the sample size. The formula for the corrected sample size is:

$$n_{adj} = \frac{n \cdot (N - n)}{N - 1}$$

Where: - **n** is the sample size without correction. - **N** is the population size.

This correction is especially useful when sampling from small populations, where not adjusting for the finite size might lead to overestimation of the required sample size.

Practical Considerations for Sample Size

- **Budget and Resources:** While a larger sample size can improve the precision of your estimates, it also increases the time, cost, and resources required for data collection and analysis. Balancing statistical precision with practical constraints is essential.
- **Sampling Method:** The sample size calculations above apply to simple random sampling. If you’re using other sampling methods (like stratified or cluster sampling), you may need to adjust the sample size accordingly, as these methods can be more or less efficient depending on the population structure.

Sample Size for Different Confidence Levels and Margins of Error

To help illustrate the effect of different margins of error and confidence levels, here’s a summary of common values for sample size calculations:

Confidence Level	Z-value	Margin of Error ± 5% (Sample Size)
90%	1.645	271
95%	1.96	384
99%	2.576	664

Conclusion

Determining an appropriate sample size is essential to ensure that survey results are reliable and precise. The sample size should be large enough to account for variability in the population while also considering practical constraints such as time, budget, and resources.