

DATA COLLECTION STRATEGIES

Quantitative Data

Amílcar Moreira

ISEG - Institute of Economics and Management
University of Lisbon

April 2, 2025

**OPEN
MINDS.
GRAB
THE FUTURE.**

Overview

1. Data Collection: Primary vs. Secondary Sources

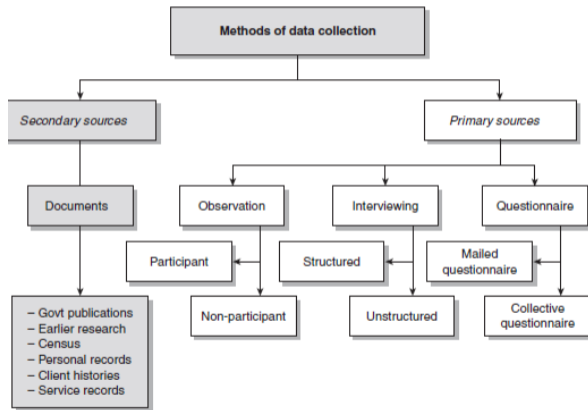
2. The Questionnaire

3. The Questionnaire: Structure, Flow, Questions

4. The Questionnaire: Sampling

Data Collection: Primary vs. Secondary Sources

Primary vs. Secondary Sources



The Questionnaire

The Questionnaire: Strengths and Weaknesses

- **Strengths**

- Stable, consistent and uniform measure, free of variation;
- Less demanding (in terms of time and resources);
- Quick results.

- **Weaknesses**

- Low response rate (as low as 20
- Self-selection bias;
- Response may be influenced by previous responses.

Source: Kumar, 2012; Sarantakos, 2012.

The Questionnaire: Modes of Administration

- **Administred by Interviewer**
 - Face-to-Face
 - Telephone
- **Self-Completion**
 - Paper Version
 - Online

Source: Bolton and Brace, 2022.

Interviewer-Administered Questionnaire

- **Advantages**

- Reduces probability of dropping-out;
- Possibility to clarify questions or to repair misunderstandings.

- **Disadvantages**

- Interviewers may fail to administer the questions exactly as they written;
- Interviewers may may record responses inaccurately;
- Increases the probability of 'Social Desirability' bias.

Source: Bolton and Brace, 2022.

Interviewer-Administered Questionnaire

- **Face-to-Face**

- Easiest to clarify questions or to repair misunderstandings;
- Strongest probability of 'Social Desirability' bias.

- **Telephone**

- Physical distance helps to reduce 'Social Desirability' bias and to discuss sensitive subjects;
- Limits ability to clarify questions or to repair misunderstandings;

Source: Sarantakos, 2012.

Self-Completion Questionnaire

- **Advantages**

- Respondents have time to consider their answers;
- Makes it easier for respondents to be honest about sensitive subjects;
- Reduces 'Social Desirability' bias;

- **Disadvantages**

- Not possible to clarify questions or to repair misunderstandings;
- Limits ability to capture immediate reactions (important in attitudinal and self-image questions).
- Increases probability of dropping-out.

Self-Completion Questionnaire

- **Paper Version**

- Typically low response rate, especially if postal return is required.

- **Online Version**

- Open Web, Email URL Embedded, Simple Email
- Platform issues (laptop, mobile, etc.)
- Using paid respondents (ex: Mechanical Turd)

Source: Bolton and Brace, 2022.

Digital Scripting Software

- **Advantages**

- Cope with complex routing;
- Rotate/Randomize/Repeat Questions/Question sets;
- Identify entry errors or consistency problems, in real time;

- **Challenges**

- Harder to check (depending on the complexity of the routing)
- May encourages 'don't know' responses (skip through),
- Less suitable for open-ended questions.

Source: Bolton and Brace, 2022.

Digital Scripting Software

- **Advantages**

- Cope with complex routing;
- Rotate/Randomize/Repeat Questions/Question sets;
- Identify entry errors or consistency problems, in real time;

- **Challenges**

- Harder to check (depending on the complexity of the routing)
- May encourages 'don't know' responses (skip through),
- Less suitable for open-ended questions.

Source: Bolton and Brace, 2022.

The Questionnaire: Structure, Flow, Questions

Questionnaire Structure

- **Introduction** (This gives respondents or potential respondents the information they require to be able to make an informed decision about whether or not they are prepared to take part in survey. Provides information on assurance of confidentiality and anonymity).
- **Body** (List of questions)

Source: Sarantakos, 2012.

Questionnaire Flow

- **Question Order Bias**
 - **Priming Effect;** Asking questions about the respondent's feelings about the issue of interest prior to the key questions can have an effect on the response to it.
 - **Consistency Effect:** Questions are structured in such a way that they feel they can only answer one way if they are to appear consistent.
- **Sensitive Questions;** May increase possibility of dropping out.

Source: Bolton and Brace, 2022.

Questionnaire Flow

- **Funnel Format**

- From general to specific, from non-sensitive to sensitive questions.

- **Inverted Funnel Format**

- From specific to general, from sensitive to non-sensitive.

- **Diamond Format**

- Combination of the inverted funnel format and the funnel format.

- **X - Format**

- The first part of the questionnaire has a funnel format and the second part an inverted funnel format.

Types of Questions

- **By Type of Response Required**

- Open or closed
- Single vs. Multiple Response Questions
- Pre-coded open questions

- **By Function**

- Exclusion / Screening questions (determine is person is (in)elegible to take part in survey)
- Filter Question (determine if the person should respond to following question / battery of questions)
- Classification Questions (self-classifications gender, age, income, final level of education, etc. Asked at the end of the questionnaire)

Scales

- **Itemized Rating Scales**
 - **Likert Scales**
- **Semantic Differential Scale**
- **Numeric Scales**

Source: Bolton and Brace, 2022.

Scales: Itemized Rating Scales

How likely are you to use the train for this journey in the near future?

- Very likely
- Quite likely
- Neither likely nor unlikely
- Quite unlikely
- Very unlikely
- Don't know

Source: Bolton and Brace, 2022: 87

Scales: Likert Scales

Do you agree or disagree with these attitudes about shopping?

	Disagree strongly	Disagree	Neither agree nor disagree	Agree	Agree strongly
Being a smart shopper is worth the extra time it takes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Which brands I buy makes little difference to me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I take advantage of special offers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I like to try new brands.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I like to shop around and look at displays.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Source: Bolton and Brace, 2022: 93

Scales: Semantic Differential Scale

How did you feel about this ad?

For each pair of statements click closest to the one that best describes how you felt about it

Worth remembering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Easy to forget
Difficult to relate to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Easy to relate to
Lively, exciting or fun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Dull
Ordinary or boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clever or imaginative
Helps to make the brand different to others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Does not make the brand any different to others
Makes me less interested in the brand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Makes me more interested in the brand

Note that the questionnaire writer alternated positive and negative ends of the scale between statements to help catch the flatliners. But dimensions three and four contain potential ambiguities.

Scales: Numeric Scales

Give *The Gingerbread Store* marks out of 100 on each of the following:
100 would mean 'perfect' and 0 'dreadful'.

The layout

Helpfulness of staff

Attractiveness of store

Source: Bolton and Brace, 2022: 99

Creating Appropriate Rating Scales

- **Word scales? Numbers? Pictures? A mix?**
- **How many scale points are required?**
- **Is a mid-point needed?**
- **Is a 'don't know' response needed?**
- **Can the scale be unbalanced, or should it have equal positive and negative points?**

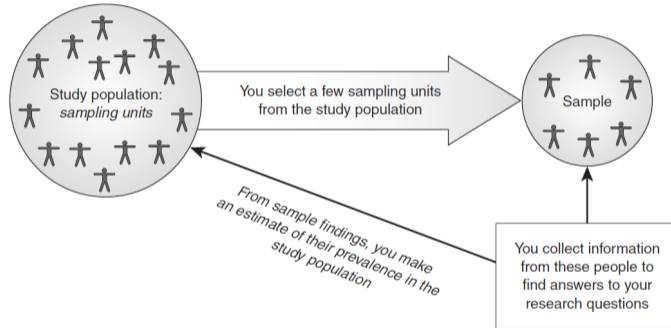
Source: Bolton and Brace, 2022.

Key Steps

1. **Preparation** (Choice of type of questionnaire and mode of administration)
2. **First Draft**
3. **External Scrutiny** (By experts)
4. **Re-Examination and Revision**
5. **Pre-Test or Pilot study** (Small sample of respondents)
6. **Revision**
7. **Second Pre-Test**
8. **Final Version**

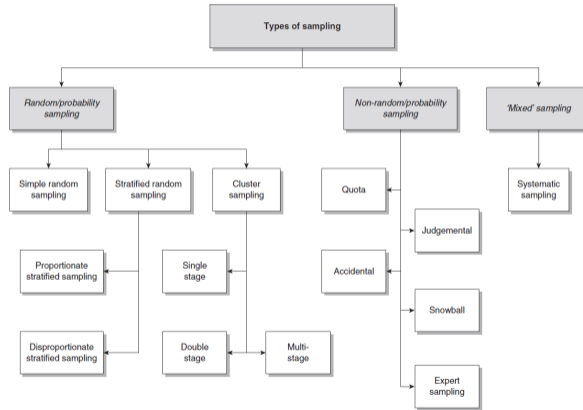
The Questionnaire: Sampling

The Concept of Sampling



Source: Kumar, 2012: 193.

Types of Sampling



Source: Kumar, 2012: 198.

Random/Probability Sampling Designs

- **Simple random sampling:** Each element in the population is given an equal and independent chance of selection. Involves pure random or systematic selection.
- **Stratified random sampling:** The population is divided into mutually exclusive strata (age, gender, etc.) that are homogeneous with respect to the characteristic on the basis of which it is being stratified. Elements are selected from each stratum, using the simple random sampling technique.
- **Cluster sampling:** Population is divided into 'clusters' on the basis of common characteristic (ex. geographical proximity). Then elements are selected within each cluster, using the simple random sampling technique. Depending on the level of clustering, sometimes sampling may be done at different levels.

Non-Random/Non-Probability Sampling Designs



- **Quota Sampling:** Selection is based on ease of access, and importance of relevant groups in the population. Respondents are recruited until the quota for each group (age, gender, etc.) is filled.
- **Purposive Sampling;** Selection is based on an assessment of who can provide the best information to achieve the objectives of the study,
- **Snowball Sampling:** Useful if you know little about the group or organisation you wish to study. Involves making contact with a few individuals, who can then direct you to the other members of the group.

Source: Kumar, 2012.

Calculation of the Sample Size

- **Rule of Thumb: *The Larger the Sample Size, the More Accurate Your Estimates***
- **What do we want from our sample?**
 - The lowest margin of error possible: 1% (0.01); 5% (0.05);
 - The lowest standar deviation possible: 0.5 is an acceptable value;
 - The highest level of confidence possible: 90%, 95% or 99

Source: Kumar, 2012.

Calculation of the Sample Size

$$\text{Sample size} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)}$$

Source:

<https://www.surveymonkey.com/mp/sample-size-calculator/>

- N = Population Size
- p = Confidence level
- z = Z-Score
- e = Margin of Error

Calculation of the Sample Size

$$\text{Sample size} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N}\right)}$$

Source:

<https://www.surveymonkey.com/mp/sample-size-calculator/>

Desired confidence level	z-score
80%	1.28
85%	1.44
90%	1.65
95%	1.96
99%	2.58

Calculation of the Sample Size

$$\text{Sample size} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N}\right)}$$

Source:

<https://www.surveymonkey.com/mp/sample-size-calculator/>

- e = Margin of Error
 - +/- 5%
 - +/- 10%



Amílcar Moreira

ISEG - Institute of Economics and Management
University of Lisbon

April 2, 2025