

# The Nature of Quantitative Research

## Chapter Summary

### Introduction

This chapter considers the overall model of quantitative research, its main steps, its main principles and goals, and also its criticisms. It shows that quantitative research is only one possible approach to studying social world, with its own preoccupations, advantages and disadvantages.

### The Main Steps in Quantitative Research

Quantitative research relies on a particular vision of the social world that allows it to propose its own model, or steps, in doing research. This vision is based on positivistic philosophy, which advocates the views that (1) social sciences may use the procedures of natural sciences in studying human world; and (2) social reality is objectively given—exists outside of our will and perceptions—and can therefore be studied (this is the objectivist view of social reality).

Because of these two main points, quantitative research proposes a top-down model of doing social research, starting from theory and moving on to collecting data and testing evidence. This model is deductive in its essence, because it moves from theory (the general) to particular (the data), and it does so by first formulating the hypotheses. Here are the main steps:

1. Theory
2. Hypothesis
3. Research design
4. Devise measures of concepts
5. Select research site(s)
6. Select research subjects/respondents
7. Administer research instruments/collect data
8. Process data
9. Findings/conclusions
10. Write up findings/conclusions

### Concepts and Their Measurement

A **concept** in this quantitative model is an idea or mental representation of something, a building block of theory represented in our research questions. Thus, in the research question “How does drug and alcohol use affects one’s self-image?” the two concepts are “drug and alcohol use” and “self-image.”

There are two definitions of any concept: (1) **nominal definition**, which describes the concept in words and is written like a dictionary entry explaining what the concept means and what

is included in it; and (2) **operational definition**, which describes how the concept will be measured or how it will actually be reflected in a question being asked of respondents. Both definitions need to be developed in a research project in order to collect and assess the evidence and test the hypotheses.

Sometimes the research considers the causal relationship between concepts, such as whether the drug use influences one's self-image. In such a situation, a concept may be seen as the cause (independent variable) or as an effect (dependent variable) that needs to be explained. It is important to note that a concept does not automatically belong in the category of causes or effects. It may be a cause in one set of research questions and an effect in another. Likewise the concept may be used simply to describe some issue or for comparison purposes rather than to establish the existence of causal relationship.

An important part of quantitative research is *to measure* concepts, in order to see how they operate in social reality. In order to be measured, the concepts must first be converted into measurable indicators, such as survey questions. The process of converting the concepts into indicators is called operationalization. Measurement in social research is useful for at least three reasons:

1. It allows us to identify fine differences between people or issues rather than fit them into broad categorizations.
2. It gives a way to identify and gauge those distinctions with consistency. We can identify whether there is a change and then work to understand what underlies such change. What characteristic changed in the research group? By establishing consistent measuring tools we are comparing apples to apples. This also relates directly to reliability issues discussed in Chapter 2.
3. Measurement allows us to estimate the relationship between concepts and the strength of that relationship. For example, we can establish whether there is a correlation between drug use and low self-image.

**Indicator** is a specific and measurable reflection of the concept in an interview question. Indicators may be formed from questions on a questionnaire or a structured interview, developed from classification criteria in observations, or adopted from official statistics or written material. Sometimes concepts are quite complex and one indicator (or one question) may not be sufficient to measure all the dimensions (parts) of the concept. Using multiple indicators to measure a single concept in a study is advantageous for several reasons:

- It reduces the likelihood of misclassifying responses when only one indicator is used and the language of that question is misunderstood.
- It ensures that all underlying meanings of the concept are captured correctly.
- It ensures that each dimension of the concept is reflected in a separate survey question, thus allowing the researcher to make finer distinctions.
- Multiple indicators also allow for sophisticated data analysis, such as factor or cluster analysis. **Factor analysis** provides a mathematical estimation of whether several selected indicators are really measuring one single concept in question. In turn, **cluster analysis** looks at whether respondents can be grouped into distinct groups based on some shared characteristics.

One very popular way to include a multiple-indicator measure in a questionnaire is through a series of Likert-type questions. These questions are usually designed to measure the intensity of a respondent's feelings about the issue, and their answer options are presented as an ordinal five-point scale from "1=strongly agree" to "5=strongly disagree," usually including a neutral position of "neither agree nor disagree" in the middle.

## Reliability and Measurement Validity

When measures (indicators) are designed, it is important to test whether they are reliable and valid representations of the concepts. **Reliability** tests the consistency of the designed measures (how stable they are) while validity tests whether they are indeed measuring the concept they claim to measure. The three most common reliability tests consider the following types of consistency:

1. *Stability over time (test–retest reliability)*: This reliability test checks that the measure does not fluctuate simply due to the passage of time and assumes that what we are measuring is not changing. This is typically conducted with a test–retest method when a test is administered to a respondent to measure a concept at one occasion ( $T_1$ ) and then re-administered to the same respondent shortly after, at  $T_2$ . If the measure of the concept is stable over time, then respondent's answers at two occasions should be highly correlated. However, stability over time is hard to ensure simply because respondent's answers at  $T_2$  may be influenced by their answer at  $T_1$  or the fact that they took a pre-test.
2. *Internal reliability (internal consistency)*: Are multiple measures administered in one sitting consistent with each other? Is the first half of indicators measuring a concept correlated with the second half of indicators measuring the same concept? This reliability test is usually conducted by the split-half method in which the first few indicators of the concept are correlated with the next batch. The two batches of indicators should show high correlation if the measure is internally consistent. The degree of this consistency is measured by *Cronbach's alpha coefficient*, and a correlation of 0.8 or higher (on a scale of 0 to 1) is generally accepted as minimum of internal reliability, although some researchers may still use lower figures.
3. *Inter-observer consistency*: All observers should classify their observations in the same way. The difficulty here is in establishing clearly defined categories and clarifying the rules of coding to reduce the possibility of variation in data interpretation. This is particularly relevant for contextualizing visual observations or interpreting open-ended questions, where researchers want to make sure that different observers code similar observations in the same way.

**Validity** ascertains whether our questions actually measure the intended concept: are we actually measuring what we want to measure? The four most common types of validity are:

1. *Face validity*: On the face of it, does it appear that the measure reflects the concept concerned? For example, measuring mathematical ability with the addition skills alone does not have face validity, because mathematical ability includes much more than addition skills. In order to determine the face validity we need to ask someone with expertise in the area to review the measures and give an opinion of whether or not the concept is reflected.
2. *Concurrent validity*: The researcher compares the measure (e.g., job satisfaction) against another criterion that is already established in the literature and is believed to be related to the concept in question (e.g., absenteeism). If the new measure is valid, the two measures are expected to correlate at a given point of time (i.e., the measures are concurrent). If such correlation is not found, this casts doubt onto the validity (correctness) of our new measure.
3. *Construct validity*: This validity test demonstrates that the concepts are related to each other in a way consistent with the researcher's theory. If the measure has construct validity, then it should be related with measures of other concepts in a way corresponding to the theory. For example, a researcher hypothesizes that happiness in marriage should be theoretically related to equal participation in household chores. If so, then a measure of participation in chores should be correlated with the measure of happiness in marriage. If such correlation is found, the measure of happiness in marriage is said to have construct validity.

4. *Convergent validity*: This type of validity gauges whether the measure of the concept developed by one method is showing the same results as the measure of the concept developed by other methods. For example, if the sociability of businesspersons found from answers to the questionnaire items corresponds to their sociability found from direct observation of their social interactions, the measure of sociability through the questionnaire is said to have convergent validity. The convergent validity is also a degree to which the measures of the concepts that are theoretically related are in fact related.

### Reflections on Reliability and Validity

In quantitative studies, addressing both reliability and validity is a sign of good research practices. Quantitative research is considered to be good if the measures it proposes are both valid and reliable. However, both reliability and validity are difficult to achieve. A measure that is not reliable will also be not valid, because unstable data cannot reflect any true concept.

However, a reliable measure can be valid in one project but not valid in another. Reliable data gathered for one project may be invalid because it does not “fit” with that particular project and measures something else. The same data might be useable in another project, reliably measuring another concept. Therefore, a measure may be reliable but not valid.

### The Main Goals of Quantitative Research

The main goals of quantitative research are as follows:

- *Measurement*: To provide data that helps to understand or quantify social phenomena, concepts, and their interrelations.
- *Establish causality*: To find causal influences on social phenomena—quantitative research does not want to simply describe how things are; it aims to explain why something occurs in society. While it is easier to see the direction of causal inference in experiments because the independent variable is manipulated, in cross-sectional designs we simply infer the direction of influence based on common sense or a prior theory. However, it is the goal of quantitative research to provide a better sense of causal direction, and various techniques such as statistical controls or longitudinal research help to improve causal interpretation.
- *Generalization*: Quantitative researchers certainly want to generalize the results of their research to a large population in order to provide the law-like explanations of social reality similar to those in natural sciences. Representative samples help to provide such generalizable explanations. Researchers use probability sampling—the procedure to draw a random sample from a given population in order to select representative samples. Random selection does not guarantee representativeness of a sample, but it does improve its chances for being representative. It must be noted that even the representative sample can only be generalized to the population it was drawn from (e.g., a sample drawn from a town only really applies to that specific town). Generalizability of research finding is of particular concern for researchers using cross-sectional and longitudinal design, while researchers in experimental camp are more concerned with internal rather than external validity.
- *Replication*: Replication is important for quantitative researchers because by making the study replicable, they minimize the researcher’s biases, values, and expectations on the outcomes. Replication, as the process of repeating a study using the same methods, provides a way to re-evaluate the findings of a research project to check for biases and routine errors, such as faulty measuring procedures and improper analysis. If the findings are not the same as those of the original study, replication provides reasons to re-evaluate the methods and findings of the original study. If the findings are the same, researchers have greater confidence in the original findings. Replication requires a detailed description of all research procedures and methods used to generate findings.

## Coding

**Coding** is a key stage in quantitative research. It has some applications in qualitative research, too, but that will be discussed in a later chapter. While coding the questionnaire, the researcher must establish categories into which the data will be sorted. For closed questions, these categories are decided upon upfront, in the process of writing the question and before the data collection. For open questions, the data are first collected and reviewed, and then the codes are developed based on the answers and themes that emerge.

When the categories in a question are established, each category is assigned a numerical value, and this numerical value is the code. Therefore, the **code** is the label given to a theme or category. For example, in the following question, “5” is the code for “*Strongly in favour*.”

*“What is your view on the legalization of marijuana? Are you strongly opposed, opposed, neither opposed nor in favour, in favour, or strongly in favour?”*

1. *Strongly opposed*
2. *Opposed*
3. *Neither opposed nor in favour*
4. *In favour*
5. *Strongly in favour*

There are some basic principles required for all coding:

- The categories must not overlap.
- The list of categories must be exhaustive, covering all possibilities. This usually requires the inclusion of a category “other.”
- Coders should be provided with a “coding frame” or “coding manual” that establishes clear rules about how codes should be applied

**Post-coding** is used with open questions and involves a review of the data for the second time, after the categories had been established, and actually assigning the code to the particular answer which shows up. A difficulty arises if the coding of the open questions is done by more than one person, since coding the verbal data might require an interpretation, and different coders may assign different codes to the same verbal answer. This inconsistency leads to measurement error and lack of validity.

**Pre-coding** requires the researcher to design a coding frame prior to administering the survey. Closed questions are by definition *pre-coded*. Each category is assigned a numerical value. The respondent is asked to select one of a set of answers that belong in the pre-set categories. With open questions the researcher allows the respondents to provide open answers, and later sorts the answers into the coding categories.

## Critiques of Quantitative Research

The most common criticisms of quantitative research are the following:

- *People and social institutions are treated as if they are part of “the world of nature.”* Critics maintain that humans and their actions are fundamentally different from the non-conscious entities that natural scientists study: subatomic particles, photosynthesis, chemical reactions, etc. Thus, the methods used to study humans must be fundamentally different from those used to study natural phenomena. Quantitative researchers counter that humans are a part of nature and that science can therefore be used to understand the human condition.
- *The measurement process produces an artificial and false sense of precision and accuracy.* Critics argue that quantitative researchers accord undue attention to measurement and wrongly assume that respondents would interpret the same survey item in the same way, thus ignor-

ing the problem of meaning. Quantitative researchers answer that their methods include tests for shared meanings, although meanings do get less attention than they deserve.

- *In its reliance on instruments and procedures, quantitative research produces a disjuncture between research and everyday life.* This is a question of validity. It is not true that survey respondents can always give knowledgeable answers to questions they encounter, or their answers may simply be different from what they do in real life. Hence, the results of the surveys are not always meaningful. Similarly, manipulation in experiments cannot copy real life and has only short-term effects. Life is a long-term, ongoing, sometimes shifting social process, and may be very different from what is recorded on a survey or what happens in an experiment.
- *The analysis of relationships between variables ignores people's everyday experiences and how they are defined and interpreted.* This incorporates the first and third criticisms above. For example, if a researcher studies the relationship between single motherhood and poverty using a questionnaire, the actual experiences of single motherhood and its meaning for the respondent will not be understood. Quantitative researchers counter that they can ask questions that get at how people interpret their everyday life.
- *Explanations for findings may not address the perceptions of the people to whom the findings purportedly pertain.* Valid quantitative findings may be achieved, but the conclusions don't take into account what the people involved actually feel and think. The categories for questionnaires are preset to analyze respondent's activity and attitudes, and they make assumptions about life that are not necessarily shared by the respondents. For example, a high number of children born to unwed parents in Western societies might lead researchers to conclude that marriage is not considered valuable in these places. In fact, qualitative research shows that this interpretation is not correct, and marriage is considered valuable to single-mothers in inner cities, but it is expected to happen later in life. Hence, the interpretation of quantitative researchers does not take subjective interpretations of participants sufficiently into account.
- *Quantitative researchers tend to assume an objectivist ontology.* Quantitative researchers tend to assume that social reality exists independently of individual will or consciousness, and they often see the social order as given, not created in interaction or negotiated. This impedes their ability of doing social research. Quantitative researchers counter that there may be certain characteristics of social reality regardless of how we perceive them, and these characteristics should be studied.

## Learning Objectives

In this chapter, you should learn to do the following:

- Understand the model of quantitative research and its main steps, which reflects a deductive (top-down) logic of research and tests a theory by gathering and analyzing data
- Formulate nominal and operational definitions of concepts and understand the difference between the two
- Appreciate the difficulties of operationalization
- List and comment on the purpose of the main goals of quantitative research: measurement, establishing causality, generalization of findings and replication
- Differentiate the main types of reliability (internal, test–retest, inter-observer) and understand their significance for replication in quantitative research

- Distinguish between the main types of validity (face, construct, convergent, and concurrent) and understand their significance for establishing causation and generalization of findings in quantitative research
- Discuss the main criticisms of the quantitative model of research

## Media Resources

Read more on Likert scaling on the following webpage:

<http://www.socialresearchmethods.net/kb/scallik.php>

- What weaknesses can you perceive with the Likert scale?
- How does use of the Likert scale respond to the criticism that the survey questions do not speak to the everyday life of people?
- Would using the Likert scale reduce the problem of misinterpretation of survey questions?
- Can the Likert scale questionnaire actually expose how people define and interpret their life experiences?
- What key process is most important when using the Likert scale questionnaire?

Refer to these two websites for more information on Cronbach's alpha coefficient:

<http://www.socialresearchmethods.net/kb/reotypes.php>

<http://www.experiment-resources.com/cronbachs-alpha.html>

- In what situations would a researcher rely on a correlation of less than 0.8?
- How does Cronbach's alpha coefficient help with internal validity?
- How does Cronbach's alpha coefficient impact generalization?

Compare Kelly and DeGraaf's (1997) and Pui-Yan Lam's (2006) cross-national research on religion. (Access to these two studies can be obtained online through JSTOR and ProQuest.

Kelley, J., & De Graaf, N. D. (1997), National Context, Parental Socialization, and Religious Belief: Results from 15 Nations, *American Sociological Review*, 62(4): 639–659. Retrieved from <http://www.jstor.org/stable/2657431>

Lam, P.-Y. (2006), Religion and Civic Culture: A Cross-National Study of Voluntary Association Membership, *Journal for the Scientific Study of Religion*, 45(2): 177–193. Retrieved from <http://www.jstor.org/stable/3838312>

- What is the impact of doing international studies?
- How does comparison of one country to another impact internal validity?
- How does comparison of one country to another impact generalization?

Review the following surveys:

General Social Survey – Victimization (GSS), Statistics Canada

<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4504&lang=en&db=imdb&adm=8&dis=2>

Uniform Crime Reporting Survey (UCR), Statistics Canada

<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3302&lang=en&db=imdb&adm=8&dis=2>

- Compare reported auto thefts to unreported auto thefts.
- What is the discrepancy?

- Why is there a discrepancy?
- What is the difference in fundamental thinking that leads to the different research methods and different results of these two types of study?