
Daniel Stockemer · Jean-Nicolas Bordeleau

Quantitative Methods for the Social Sciences

A Practical Introduction with Examples
in R

Second Edition

Contents

1	Introduction	1
2	The Nuts and Bolts of Empirical Social Science	5
2.1	What is Empirical Research in the Social Sciences?	5
2.2	Qualitative and Quantitative Research	8
2.3	Theories, Concepts, Variables, and Hypotheses	10
2.3.1	Theories	10
2.3.2	Concepts	12
2.3.3	Variables	13
2.3.4	Hypothesis	16
2.4	The Quantitative Research Process	18
	References	20
3	A Short Introduction to Survey Research	23
3.1	What is Survey Research?	23
3.2	A Short History of Survey Research	24
3.3	The Importance of Survey Research in the Social Sciences and Beyond	26
3.4	Overview of Some of the Most Widely Used Surveys in the Social Sciences	27
3.4.1	The Comparative Study of Electoral Systems (CSES)	28
3.4.2	The World Value Survey (WVS)	29
3.4.3	The European Social Survey (ESS)	29
3.5	Different Types of Surveys	30
3.5.1	Cross-Sectional Survey	30
3.5.2	Longitudinal Survey	32
	References	34
4	Constructing a Survey	37
4.1	Types of Questions a Researcher Can Ask	37
4.2	Ordering of Questions	38
4.3	Number of Questions	38
4.4	Getting the Questions Right	39
4.5	Social Desirability	41

4.6	Open-Ended and Closed-Ended Questions	42
4.7	Types of Closed-Ended Survey Questions	44
4.7.1	Scales	44
4.7.2	Dichotomous Survey Question	47
4.7.3	Multiple-Choice Questions	47
4.7.4	Numerical Continuous Questions	48
4.7.5	Categorical Survey Questions	48
4.7.6	Rank Order Questions	49
4.7.7	Matrix Table Questions	49
4.8	Different Variables	50
4.9	Coding of Different Variables in a Dataset	51
4.9.1	Coding of Nominal Variables	52
4.10	Drafting a Questionnaire: General Information	52
4.10.1	Drafting a Questionnaire: Step-By-Step Approach	53
4.11	Example of Questionnaire	54
4.11.1	Background Information About the Questionnaire	55
	References	56
5	Conducting a Survey	59
5.1	Population and Sample	59
5.2	Representative, Random, and Biased Samples	60
5.3	Sampling Error	63
5.4	Non-random Sampling Techniques	64
5.5	Different Types of Surveys	66
5.6	Which Type of Survey Should Researchers Use?	68
5.7	Pre-tests	69
5.7.1	What is a Pre-test?	69
5.7.2	How to Conduct a Pre-test?	70
	References	71
6	Introducing R and Univariate Statistics	73
6.1	R Programming Language	73
6.1.1	Downloading R and RStudio	73
6.1.2	RStudio Interface	74
6.1.3	R Packages	75
6.1.4	The Basics of R	76
6.2	Importing Data into R	77
6.3	Frequency Table	79
6.3.1	Constructing a Frequency Table in R	79
6.4	Measures of Central Tendency	80
6.4.1	Mean	80
6.4.2	Median	81
6.4.3	Mode	81
6.4.4	Range	81
6.4.5	Measures of Central Tendency in R	81

6.5	Displaying Data Graphically with Pie Charts, Boxplots, and Histograms	82
6.5.1	Pie Charts	82
6.5.2	Boxplot	84
6.5.3	Histogram	85
6.6	Measures of Dispersion, Sampling Error, and Confidence Intervals	86
6.6.1	Calculating Confidence Intervals in R	90
	References	90
7	Bivariate Statistics with Categorical Variables	93
7.1	Independent Samples t-Test	93
7.1.1	Calculating a t-Value for Independent Samples t-Test	95
7.1.2	Doing an Independent Samples t-Test in R	96
7.1.3	Interpreting an Independent Samples t-test	98
7.1.4	Reporting the Results of Our Independent Samples t-test	98
7.2	One-Way Analysis of Variance (ANOVA)	99
7.2.1	One-Way Analysis of Variance in R	100
7.2.2	Interpreting the Results of an ANOVA	101
7.2.3	Post-hoc or Multiple Comparison Tests in R	101
7.2.4	Reporting the Results of an ANOVA and Post-hoc Comparison Tests	103
7.3	Cross-Tabulation Tables and Chi-Square Test	103
7.3.1	Cross-Tabulation Tables	103
7.3.2	Chi-Square Test of Independence	105
7.3.3	Chi-Square Tests in R	106
7.3.4	Interpreting a Chi-Square Test Conducted in R	107
7.3.5	Reporting the Results of a Chi-Square Test	107
	References	107
8	Bivariate Statistics with Two Continuous Variables	109
8.1	What is a Bivariate Relationship Between Two Continuous Variables?	109
8.1.1	Positive and Negative Relationships	109
8.2	Scatterplot	110
8.3	Positive Relationship Displayed in a Scatterplot	110
8.4	Negative Relationship Displayed in a Scatterplot	110
8.5	No Relationship Displayed in a Scatterplot	110
8.6	Drawing a Line in a Scatterplot	112
8.7	Building a Scatterplot in R	112
8.8	Correlation Analysis	114
8.9	Computing a Correlation Analysis in R	117
8.9.1	Interpreting and Reporting the Results of a Correlation Using R	118

8.10	Bivariate Regression Analysis	118
8.11	Gauging the Steepness of a Regression Line	118
8.12	Gauging the Error Term	120
8.13	Computing a Bivariate Regression Analysis in R	122
8.14	Interpreting the Regression Output	123
8.14.1	Regression Coefficient and Intercept Estimates	123
8.15	Standard Error and t-Value	123
8.16	Model Fit	123
8.17	Reporting Regression Results with a Model Table	125
8.18	Presenting the Results in a Research Article	125
	References	126
9	Multivariate Regression Analysis	127
9.1	The Forms of Independent Variables to Include into Multivariate Regression Models	129
9.2	Interpreting a Multivariate Regression Model	129
9.3	Computing a Multiple Regression Model in R	130
9.4	Interpreting a Multiple Regression Model	131
9.5	Reporting the Results of a Multiple Regression Analysis	132
9.6	Finding the Best Model	132
9.7	Assumptions of the Ordinary Least Squares Regression Model (OLS)	133
	References	136
	Appendix 1: The Data of the Sample Questionnaire	137
	Appendix 2: Possible Group Assignments that Go with This Course	139
	Book Summary	141



Introduction

1

Under what conditions do countries go to war? What is the influence of the 2008–2009 economic crisis on the vote share of radical right-wing parties in Western Europe? What type of people are the most likely to protest and partake in demonstrations? How has the urban squatters’ movement developed in South Africa after Apartheid? What is the impact of the COVID-19 pandemic on conspiracy beliefs? There is hardly any field in the social sciences that asks as many research questions as political science. Questions scholars are interested in can be specific and reduced to one event (e.g., the development of the urban squatter’s movement in South Africa post-Apartheid) or general and systemic such as the occurrence of war and peace. Whether general or specific, what all empirical research questions have in common is the necessity to use adequate research methods to answer them. For example, to effectively evaluate the influence of the economic downturn in 2008/2009 on the radical right-wing success in the elections preceding this crisis, we need data on the radical right-wing vote before and after the crisis, a clearly defined operationalization of the crisis, and data on confounding factors such as immigration, crime, and corruption. Through appropriate modeling techniques (i.e., multiple regression analysis on macro-level data), we can then assess the absolute and relative influence of the economic crisis on the radical right-wing vote share.

Research methods are the “bread and butter” of empirical political science. They are the tools that allow researchers to conduct research and detect empirical regularities, causal chains, and explanations of political and social phenomena. To use a practical analogy: a political scientist needs to have a toolkit of research methods at his or her disposal to build good empirical research in the same way as a mason must have certain tools to build a house. It is indispensable for a mason to not only have some rather simple tools (e.g., a hammer), but also some more

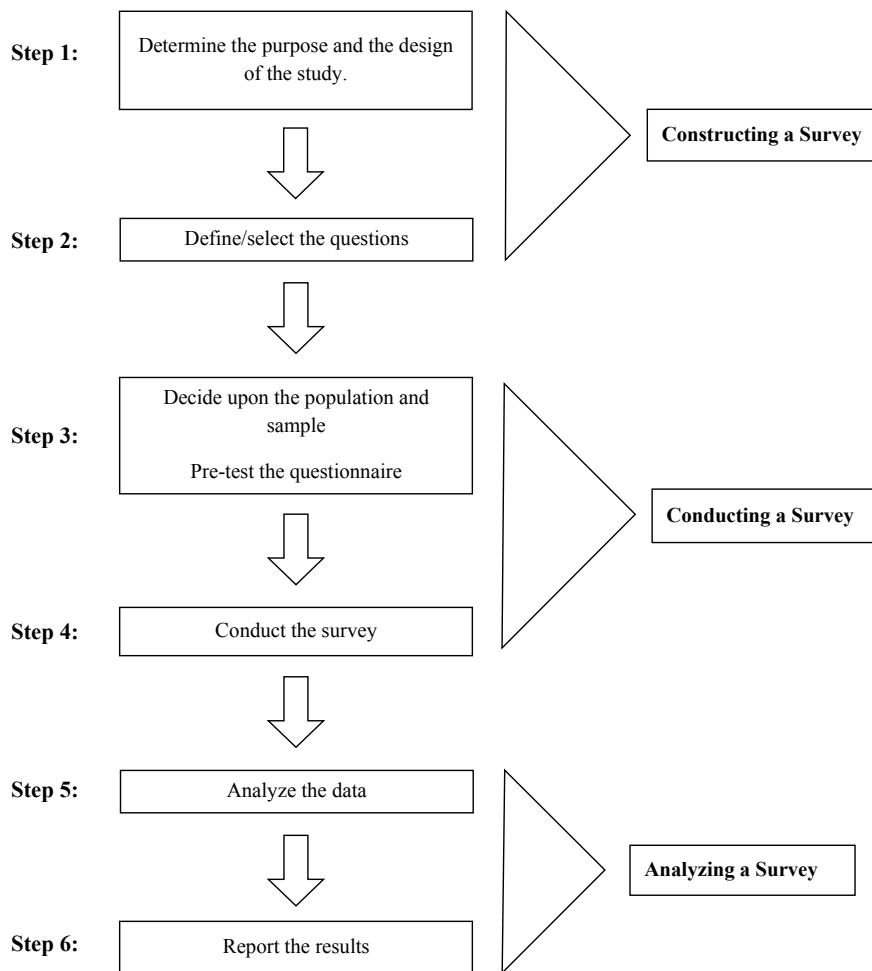


Fig. 1.1 Different steps in survey research

measures the amount of money students spend partying. It is an original survey including the original data, which a previous student group collected during their semester-long project. Using this “colloquial” survey, the students in this study group had lots of fun collecting and analyzing their “own” data, showing that learning statistics can (and should) be fun. We hope that the readers and users of this book experience the same joy in their first encounter with quantitative methods.



The Nuts and Bolts of Empirical Social Science

2

2.1 What is Empirical Research in the Social Sciences?

Regardless of its subdiscipline, empirical research in the social sciences tries to decipher how the world works around us. Be it development studies, economics, sociology, political science, or geography, just to name a few disciplines, researchers try to explain how some part of the world is structured. For example, political scientists may try to answer why some people vote while others abstain from casting a ballot. Scholars in developmental studies might look at the influence of foreign aid on economic growth in the receiving country. Researchers in the field of education studies might examine how the size of a school class impacts the learning outcomes of high school students, and economists might be interested in the effect of raising the minimum wage on job growth. Regardless of the discipline they are in, social science researchers try to explain the behavior of individuals such as voters, protesters, students, the behavior of groups such as political parties, companies, or social movement organizations, or the behavior of macro-level units such as countries.

While the tools taught in this book are applicable to all social science disciplines, we mainly cover examples from empirical political science, because this is the discipline in which we teach and research. In all social sciences and in political science, more generally, knowledge acquisition can be both normative and empirical. Normative political science asks the question of how the world ought to be. For example, normative democratic theorists quibble with the question of what a democracy should be. Is it an entity that allows free, fair, and regular elections, which in the democratic literature, is referred to as the “minimum definition of democracy”? (Boogards, 2007). Or must a country, in addition to having a fair electoral process, grant a variety of political rights (e.g., freedom of religion, the freedom of assembly), social rights (e.g., the right to health care and housing), and economic rights (e.g., the right to education or housing) to be “truly” democratic? This more encompassing definition is currently referred

to in the literature as the “maximum definition of democracy” (Beetham, 1999). While normative and empirically oriented research have fundamentally different goals, they are nevertheless complementary. To highlight, an empirical democracy researcher must have a benchmark when she defines and codes a country as a democracy or non-democracy. This benchmark can only be established through normative means. Normative political science must establish the “gold-standard” against which empirically oriented political scientists can test whether a country is a democracy or not.

As such, empirical political science is less interested in what a democracy should be, but rather how a democracy behaves in the real world. For instance, an empirical researcher could ask the following questions: Do democracies have greater gender representation in parliaments than non-democracies? Do democracies have less military spending than autocracies or hybrid regimes? Is the history curriculum in high schools different in democracies than in other regimes? Does a democracy spend more on social services than an autocracy? Answering these questions requires observation and empirical data. Whether it is collected at the individual level through interviews or surveys, at the meso-level through, for example, membership data of parties or social movements, or at the macro-level through government/international agencies or statistical offices, the collected data should be of high quality. Ideally, the measurement and data collection process of any study should be clearly laid down by the researcher, so that others can replicate the same study. After all, it is our goal to gain intersubjective knowledge. Intersubjective means that if two individuals would engage in the same data collection process and would conduct the same empirical study, their results would be analogous. To be as intersubjective or “facts based” as possible empirical political science should abide by the following criteria:

Falsifiability: the falsifiability paradigm implies that statements or hypotheses can be proven or refuted. For example, the statement that democracies do not go to war with each other can be tested empirically. After defining what a war and democracy is, we can get data that fits our definition for a country’s regime type from a trusted source like the *Polity IV* data verse and data for conflict/war from another high-quality source such as the *UCDP/PRIO Armed Conflict* dataset. In second step, we can then use statistics to test whether the statement that democracies refrain from engaging in warfare with each other is true or not.^{1,2}

Transmissibility: the process through which we can achieve the transmissibility of research findings is called replication. Replication refers to the process by which prior findings can be retested. Retesting can involve either the same data or new

¹ The PolityIV database adheres to rather minimal definition of democracy. In essence, the database gauges the fairness, competitiveness of the elections and the electoral process on a scale from –10 to +10. –10 describes the “worst” autocracy, while 10 describes a country full respects free, fair, and competitive elections.

² The UCDP/ PRIO Armed Conflict Dataset defines minor wars by a death toll between 25 and 1000 people and major wars by a death toll of 1000 people and above (see Gleditsch, 2002).