POLITICAL ANALYSIS

Matthew Loveless

POLITICAL ANALYSIS

A Guide to Data & Statistics

Second Edition

Sage

Sage

1 Oliver's Yard
55 City Road
London EC1Y 1SP

2455 Teller Road
Thousand Oaks
California 91320

Unit No 323-333, Third Floor, F-Block
International Trade Tower
Nehru Place, New Delhi – 110 019

8 Marina View Suite 43-053
Asia Square Tower 1
Singapore 018960

Editor: Andrew Malvern
Editorial assistant: Clara Landgren
Production editor: Martin Fox
Copyeditor: Richard Walshe
Proofreader:
Indexer:
Marketing manager:
Cover design:
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed in the UK

Library of Congress Control Number: 2025939864

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-03-621693-1
ISBN 978-1-03-621692-4 (pbk)

Contents

Online resources

This second edition of *Political Analysis* is supported by a wealth of online resources for instructors to support teaching, which are available at: https://study.sagepub.com/loveless2e

For instructors

Teaching guides outline the key learning objectives covered in each chapter and provide suggested activities/examples to use in class or for assignments.

PowerPoint decks featuring figures and tables from the book, which can be downloaded and customized for use in your own presentations.

Workbooks containing questions related to the key concepts in each chapter can be downloaded and/or assigned to students and used in class, as homework or in exams. These come with solutions in Stata, SPSS, R, and Python.

Datasets for you to share with your students in class or for assignments, which will support their mastery of data analysis and statistics.

About the author

Matthew Loveless is an Associate Professor in the Department of Political and Social Sciences at the University of Bologna (Italy). He is also co-founder of the Center for Research and Social Progress (cersp.org). He has taught quantitative methods to undergraduate and graduate students since 2003. He has held academic positions in the United States (Georgetown University; University of Mississippi), the United Kingdom (Nuffield Fellow, Oxford; University of Kent), and Italy (Jean Monnet Fellow, European University Institute, Florence; University of Bologna) in addition to visiting positions at Sciences Po – Institut d'Etudes Politiques de Grenoble (France), the University of Georgia (USA), Davidson College (USA), St. Antony's College (Oxford, UK), Mannheimer Zentrum für Europäische Sozialforschung (Germany), and the University of Debrecen (Hungary). His research interests include the field of Political Behaviour in Europe, particularly as it relates to how individuals perceive and understand politics (recent examples focusing on political attitudes include *International Political Science Review*, *Political Studies*, *the Journal of European Public Policy*). Recent publications also include co-authored work that incorporate party competition in *Government and Opposition, Electoral Studies,* and the *Journal of Common Market Studies*. He lives with his family in Italy.

Prologue

This book is only nominally about statistics. It is substantively about empowering you. Learning statistics is like learning any skill. While statistical skills can help you with the consumption and execution of quantitative analysis, they can also equip you with the power of logical thinking, thoughtful analysis, and proper interpretation. That is, the practice of critical and clear thinking can be used as a shield and sword against unclear logic, specious arguments, and deception. I want you to have the confidence to interact with statistics and, in doing so, make you a more formidable student and citizen. During your study, if you start losing motivation or feeling overwhelmed, return to this paragraph to remind yourself of the real value of being an educated person.

Introduction

The aim of this book

This book was designed to help you pass through unfamiliar territory. This is one of the reasons 'guide' is in the title. However, your destination – the point of the journey – can be any number of things. It could be simply to pass this class, move on, and put it behind you. Or it could be gaining a sense of what statistics can do – and not do – so that you are not excluded from areas of research and debate. Or it could be the first of many steps toward a more advanced study of statistics and other quantitative methods. Each of these is perfectly understandable and reasonable to the extent it is appropriate to your own direction. I hope this book helps you achieve your goal.

My aim with this book is to help you to understand – before the numbers and formulas and computations – *what it is we want to know* and *why we want to know it*. Then, the computation of *how we go about finding it* becomes more intuitive and practical. I would like for you to understand the basic elements of a statistical approach to political analysis in an intuitive and useful way.

Alan Watts was a student of philosophy and attracted substantial attention in the 1960s and 1970s as an interlocutor of Eastern religions and philosophies for Western audiences. His unique appeal was the ability to offer an attainable appreciation of the context and mind necessary to comprehend – and thus begin to enjoy – oriental philosophies, ideas, and approaches. He argued that there are two sides of the mind that must move together, the intellectual mind – to understand and remember ideas correctly – and the intuitive mind – knowing when and how those ideas are appropriate. His superficially humorous but fundamental command to audience members was to 'get out of your own way!'

This has clear analogue to learning statistics. Statistics can be technically difficult and feel foreign. However, if we can understand what we are doing with data – to know our goal, to know what want to know from these data and our analysis of it, and to know why it is important to do so – we can see beyond the formulas and calculations. Such considerations take place long before we sit down with the data, and even beyond the sphere of statistics itself. This is why most statistics textbooks start off – as I do here – with talking about their use in the broader context of political analysis and the scientific method. Thus, the art of science, being a craftsman of science, is knowing when to reach for the tool of statistics as the best means to analyse a problem, and when to leave it on the workbench. Simply using a hammer is not hard: bang, bang, bang. Knowing when best to use a hammer is the trick.

Why study statistics at all?

Whether you just want to be done with a statistics class, sated with this much statistics, or are thirsting for more, there are three really good reasons to study statistics: statistical literacy, statistical abilities, and research skills.

Statistical literacy

Statistical literacy is reading, understanding, and critically assessing quantitative research. I would argue that whether you like statistics (or see some value in having been exposed to it) or not, you benefit from a statistics course. If you like statistics ('statistics-curious,' perhaps), understanding core concepts that underpin the modern use of statistics is in itself valuable and serves as a solid basis for exploring more advanced applications.

If you don't like statistics, then this is the opportunity to get to know your enemy. If you think that statistics are a waste of time or, worse, obscuring real relationships and findings, understanding where statistics does in fact stand on thin ice or rely on strong assumptions can only empower your position. That is, instead of saying, 'You're wrong to use statistics,' it would be more powerful to say, 'You're wrong to use statistics *because*…'. Learn statistics to be a better critic – as throwing uninformed insults is gauche and lazy. Or in a phrase loosely attributed Pablo Picasso, 'Learn the rules like a pro, so you can break them like an artist.'

In any case, both groups can further benefit – in different ways – from understanding what statistics can do as well as understanding what statistics cannot do.

Another reason is that a lot of what we know about the world is discovered and often explained in the language of science and the use of statistics. A great deal of what you are hearing about in your other classes comes from the cumulative efforts of scientists using various tools to advance our collective knowledge.

Consider when your other professors say:

- Citizens with socio-economic characteristics A, B, and C in institutional contexts of Y and Z are more likely to vote.
- Initial conflicts with characteristics D, E, and F in regions S and T are more likely to escalate in violence.
- Countries with characteristics M and N are more likely to be democracies.

These results are summarizing predominantly scientific – and very often statistical – analyses. These patterns are derived from accumulated empirical studies and represent conclusions based on those findings. Simply, a lot of what we know about politics, about the world more generally, is the summary of scientific research. And a lot of that research uses statistics.

And, regardless of what you do next, you will be confronted by statistics. This can be in the garden variety form of the news reporting on a legislative outcome, an election, the level of support in an article online, or even in advertising. Or perhaps you will be confronted by some research, a policy paper, white paper, or report for your work. Which position would you rather be in: passing it to a colleague because you can't understand it or engaging with it, asking questions like:

- How appropriate are the data to answering the question?
- Are the chosen statistical techniques appropriate?
- What assumptions have been made?
- What can we conclude about the empirical relationships?
- Have the limitations been acknowledged and even addressed?

This would make you a formidable employee and, more crucially, a formidable democratic citizen. Thus, statistical literacy is not about running around and analysing everything statistically, it is about not being pushed around.

Why let other people explain the world to you?

Statistical abilities

Statistical abilities include generating descriptive and inferential statistics using statistical software and interpreting and analysing the output. Not only can being statistically literate allow us to identify and use the most appropriate techniques as well as to recognize inherent limitations, it can help immunize us from bad arguments, questionable correlations, and poor inference.

As we will see, the most advanced techniques are reconfigurations of a small set of fundamental concepts. Once you digest those fundamentals, the rest will flow naturally. The challenge of statistics is learning which technique to use and how to interpret the results. Statistical abilities are recognizing the most effective means to represent relationships among data; to understand the utility of statistics ('What we can do with statistics') and the limitations of statistics ('What we can't do with statistics').

The value of a good education is not simply being exposed to more books or better teaching; a good education is becoming responsible for your own learning. And in precisely this way, the student becomes the master.

Research skills

Research skills are the rigorous use of statistics to address a research question and present your choices and results in a clear, informative, and effective manner. Simply, to use statistical techniques to inform and/or support an argument.

It may be becoming increasingly apparent to you that fewer and fewer people – certainly ones that we don't know personally – care very little about what you think, or what you feel, or what you believe. That is not to say that these are not important to you as a person. However, in writing persuasively and convincingly, people are primarily moved by what you can demonstrate. In contrast to opinions and assertions *ipse dixit* ('he asserts without proof'), statistical abilities can be used to support and advance an argument. Statistics are not a replacement for an argument but a powerful analytical tool grounded in the scientific method.

Research skills also inform our choice to use statistics. A growing number of users of statistics are charging into any question with more and more sophisticated statistical techniques. They have not received their statistical training embedded in the crucial context of scientific research and merely wield a heavier hammer. Which, despite its intimidating size and terrifying power, *still* does not make everything a nail.

Research skills are understanding how to ask a valuable question and how to bring the evidence necessary to resolve that question. In this way, we are better served by people who can understand what can be done with a statistical approach as well as what can't be done to produce the research we want – and the research we need.

Math skills

Don't be intimidated by the formulas, and letters, and mathematics as statistical software has removed the challenge of calculation. It may surprise you, but if you can do the following, you have more than sufficient math skills to do statistics:

$3 + 2 = 5$

$8 - 5 = 3$

$4 * 5 = 20$

$30 \div 10 = 3$

$8.1 > 7.4 = \text{true}$

The data in this book

One choice that was important in creating this book was making the conscious and time-consuming decision to use actual, publicly available data. Except in a few cases, the data – and subsequent results – in the exercises and examples are available to you. Available to anybody.

The primary reason is that I want you to do it. I want you to be able to use what you are learning on real data so that you can see that there is no hokey-pokey going on. You can answer and address real questions, to examine real people and events. I want you to stay close to an actual analysis of reality.

I will admit that it was tempting to include constructed examples that would perform precisely and cleanly. However, the cloudy answers the academics and researchers are familiar with are the norm. It is uncommon to find convincing, overwhelming, and clear-cut evidence. Sometimes, (statistical) results are not obviously one thing or another. Or worse, the nagging feeling that while the answer may be sufficient, it is not satisfying.

For the open-minded and patient scientist, however, these cloudy answers are not necessarily a setback but rather an opportunity, an opportunity to reflect on the nature of the results. Are we asking the right question? Perhaps the results are cloudy and underwhelming because we have approached our phenomenon of interest hindered by unclear or lazy thinking, scant understanding of previous work, or with too tight a tether to these same previous works. If not our question, perhaps it is our method of inquiry.

These are not statistical questions *per se* or even scientific questions. They are broader, more philosophical questions to improve the likelihood that you will find an answer your research. That is to say, be prepared to reflect on what it is you have really found statistically, empirically, epistemologically, even ontologically.

Finally, pedagogically, from my two decades of teaching statistics, I want you to use real data, because why tinker around the edges of real political analysis? Pick it up and do it.

Take a friend: companion workbooks in spss, stata, r, and python

As a guide for using statistics in Political Science, Political Analysis can be used on its own (a vade mecum, if you will). However, students can supplement the topics and techniques with a workbook that provides both practical and theoretical examples using different statistical software, including SPSS, Stata, R, or Python: The SPSS Workbook for Political Analysis, The Stata Workbook for Political Analysis, The R Workbook for Political Analysis, and The Python Workbook for Political Analysis.

This workbook is the wingman to your textbook. Each Workbook gives students the opportunity to explore – chapter by chapter – the topics and techniques discussed in Political Analysis as well as develop your use of statistical software to produce descriptive statistics, correlation, inferential statistics, regression, and logistic regression. The exercises in the workbooks expand on textbook examples as well as provide sets of new questions. The included datasets can reproduce many of the examples in the textbook and include new variables so that students can develop existing knowledge and practice.

The datasets are derived from the European Social Survey, Quality of Government, V-Dem, and World Values Surveys, and others. The workbooks also cover other topics including an introduction to statistical software, using publicly available datasets, cleaning and transforming data for use, producing original research, and reminders on the ethics of data handling. The solutions – answers to all the questions and the code needed to produce the output – are available to all textbook adopters.

Statistics are a tool

Data can become information. Information can then become knowledge.

Converting data into information is assembling observations into a coherent form. Statistics are one of many ways of doing this. However, statistics are not the answer. They are not the question. They are simply a tool. In their service to science, statistics are popular because they provide a straight-out-of-the-box utility.

Keep in mind that statistics can only help us with what we can measure, and data are nearly always incomplete. Incomplete does not mean lacking (although this is also a problem) but rather the gap between what we want to measure - and what we have to measure it.

Think about a person that you kind of know, a friend of a friend. How would you quantify the following attributes: age, income, gender, vote in last election, political participation, trust in government, ideology? How would you measure and record these? What if you had to collect the same data on 10 other people, 100 other people, a thousand, ten thousand? Would this change your strategy in collecting these data? In trying to find some way to compare large numbers of things, we are forced to make some concessions. That is, at some point we make our peace with imperfect measures in order to get, however imprecisely, at 'something'.

This is the first partition from reality.

The second partition is that statistics are just a way that we have developed to identify patterns in data. We have done so by exploiting mathematics (which is admittedly a very good way at identifying patterns). Werner Heisenberg – yeah, that guy – noted that 'what we observe is not nature in itself but nature exposed to our method of questioning' (2007: 24–5). If we ask statistical questions, we will get statistical answers. That is, if we seek to identify patterns, we will get answers in the form of patterns.

What we want to do is to be able to identify patterns in a sensible way so as to strengthen our understanding of the world. Yet, patterns are summaries. Some of them will provide more evidence, some less evidence. Some will be complex and some will be less complex. Yet, in the same way that we come to take for granted that there is a constellation called Big Bear (*Ursa Major*) in the distribution of stars in the sky, a pattern that fits our expectation of a pattern that should be there doesn't necessarily mean that it is actually there (Pearl and Mackenzie, 2018). That is, understanding these patterns become how we understand something – but the patterns are neither the *something* nor *understanding* itself. That comes from a different process of turning information into knowledge.

As the old joke goes, information is knowing that a tomato is a fruit, and knowledge is not putting a tomato in fruit salad. In precisely the same way, statistics are a cold, inert tool. We must separate what statistics can do for our analysis as well as what they cannot do, such as come up with good research questions, survey the literature, derive appropriate hypotheses, or correctly interpret the results.

The English figure of speech 'a jack of all trades, master of none' is often meant derogatorily to describe someone good at several things but not great at any one thing. However, this phrase originally had the opposite intention, as the full quote is: 'A jack of all trades, master of none is oftentimes better than a master of one.' That is, expertise in only one area can also lead us to view all problems through that single lens.

In this way, the study and use of statistics to analyse data as part of the larger scientific approach can go well beyond merely using statistics and serve you well in any situation that requires rigorous, analytical thinking.

A parting note

I hope that you will discover that even though having written a textbook on statistics, I am not dogmatic about it. Statistics are an analytical research tool. Common, but one of many empirical techniques. Science is a framework of understanding. Common, but one of many branches of philosophy. And both, often found together, provoke much larger philosophical questions, ontological questions, epistemological questions. If this is interesting to you, and I hope that it is, ask your instructor for further discussions on these topics.

Acknowledgements

I will assume by the fact that you are reading this sentence that you have to take a statistics course (or you have a strange sense of what fun is). Well, things could be worse. You could have been asked to write a statistics textbook. So, swings and roundabouts, silver linings and all that. It is profoundly true that a book is never written alone. I want to deeply thank the team at Sage, in particular Andrew Malvern and Clara Landgren. I am also indebted to Simon Luck for his help in developing the code for the workbook. All errors remain my own.

Finally, my greatest appreciation goes to my family. *Amore mio, non sono sicuro di poterti ringraziare abbastanza per il tuo aiuto, il tuo contributo e il tuo sostegno.* For you other three, I hope you enjoy getting a statistics textbook for your birthdays this year (again!). To put us in the right mindset for acquiring a new skill, I include a quote I once heard only once from an unseen speaker many years ago: 'Catch the vigorous horse of your mind.'

And here we go.

Annotated references and further reading

King, Gary. 1986. 'How not to lie with statistics: Avoiding common mistakes in quantitative political science' *American Political Science Review* 30(3): 666–687.

A helpful guide by one of the premier methodologists in Political Science.

Pearl, Judea and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Data alone are hardly a science, regardless how big they get and how skilfully they are manipulated.

Heisenberg, Werner. 1958. *Physics and Philosophy: The Revolution in Modern Science*. New York: Harper.

Heisenberg provides lessons for scientists of all stripes in our shared form of thinking as well as its undeniable shortcomings common across all fields of inquiry. A seminal work of scientific thought.

3

Data and variables

Learning outcomes

By the end of this chapter, you will be able to

- Describe the importance of data to Political Science research.
- Critically assess the relationship between data and variables.
- Identify the crucial importance of the conceptual basis of variables.
- Evaluate the process of operationalizing variables.
- Articulate the challenges to the choice of indicators for variables.

Introduction

Data and variables are the building blocks of scientific research. How we define the elements of our research and select the measures – or 'indicators' – of those concepts are often overlooked steps in the research process. While theory may guide our choices, what to include and exclude in our research will shape our results in important ways.

Data

There is a lot of discussion about data these days. Big data, the availability of data, data collection, data on elections, data on GDP *per capita* growth, download the data, send me the data, data on manifestos, data on conflict, here's the data, data, data, data!

So, what are data? Data, a collection of singular **datum** combined in the form of a **dataset**, are merely systematically collected, codified observations. Codified observations are observations rendered mathematically tractable or, at a minimum, discretely categorized. 'Discretely categorized' is being able to separate observations into a set of comprehensive and exclusive categories.

Frankly, what data *are* is far less important than what data can *do* for scientific research. Data are inert. Left alone, they will literally just sit there taking up space on your hard drive or cloud. So, how do data become useful to research? The purpose of data to statistical analysis is to provide observations on all the relevant elements to help determine if one explanation is better than another. That is, data become useful when we press them into the service of our research.

Data provide us with the material for the building blocks of scientific research: samples and variables. Samples allow us to efficiently study large and often distant processes in order to connect the ideas under investigation with the 'real world'. Data render research ideas, theories, and hypotheses into things that can be measured, compared, and tested. They are the proverbial rubber that meets the road. Thus, when we talk about data, we are talking about the elementary components of all research. Data are the quarks to matter, the clay to bricks, the grain to pasta.

In order for the use of statistics to have any power at all, we must consider two things. First, we must consider the actual techniques in statistics, which is what the vast majority of any book on statistics is about: the selection of the most appropriate techniques, the formulas, and drawing conclusions. We will get to this shortly.

The other part to consider is the grist for the (statistics) mill. Statistical techniques can do myriad simple and amazingly complex transformations but if you feed them slop, you will get slop. Simply, the most sophisticated mathematical handling of data is moot if in fact the data are garbage.

GIGO

Spanning back to the earliest days of computer programming (the 1950s!), the 'GIGO' concept – garbage in, garbage out – is the origin of this sentiment. Simply, no matter the technique, the quality of output is highly contingent on the quality of the input.

So, let's look at how data become useable to the statistical techniques we will learn later in the book.

For any research project, the quality of data is of utmost importance. Assuming that the data come from a reliable source and is sufficient, e.g., that there is enough of it to answer our question and control for others' explanations, data are transformed into elements of scientific research if they meet the following two demanding criteria when pressed into service. Data must:

1 be representative of what we are trying to study
2 be what we want to measure

One: Data must be representative of what we are trying to study

Data must be representative of the population they are meant to refer to. If you are interested in Southeast Asian voters, we need data representative of Southeast Asian voters. If you are interested in conflicts in sub-Saharan Africa, you need data representative of conflicts in sub-Saharan Africa. If you are interested in environmental legislation in the European Union Parliament, you need data representative of environmental legislation in the European Union Parliament. *Ad infinitum*. In order to achieve our scientific goal of inference, we need a valid – i.e., representative and sufficient – sample of the unobservable total population that we are interested in. So, let's talk about samples and populations.

Samples and populations

A **population** is the entire set of what you wish to draw conclusions about. If you are interested in why some citizens vote for Green parties or radical right parties in a country (or a region, like Europe), it would be very difficult to ask everyone in that country (or region) in a systematic and timely manner about their support or vote for these parties. In order to understand such a phenomenon, a researcher can instead take a **sample** of the population in which s/he is interested.

A sample is a subset of units in the population of interest. This is not laziness, this is necessity. In other words, because it is nearly always impossible to collect data on an entire population simultaneously, we take a sample of that population on which to conduct our investigation. If the sample is valid (that is, it 'looks like' the population from which it was drawn), the conclusions that we reach can be exported to our understanding of the population from which the sample was drawn. The degree of confidence we can have exporting the findings from the sample to the population will be given to us by the statistical techniques we will learn. Note that while correlation is more well-known, the unique superpower of statistics is conferring the possibility of making confident inferences from samples to populations.

Clearly, samples – valid and sufficient samples – are crucial to our ability to make confident inferences. While sampling procedures vary between how we conduct surveys or collect data on countries, one thing underpins the collection process: randomization. Our ability to make confident inference is made possible by the assumption that we have a random sample of the population (technically this assumption allows us to estimate the probability that a sample result could be due to chance). With randomized samples, inferential statistics will tell us *exactly* how confident we can be to make such claims. At this point, you can see how this ability might be useful to scientific research (think Principle 3).

We will take up several of these concepts – such as randomization and inference – in later chapters. For now, samples are collections of data that hinge on their representativeness – often assumed from random selection – of the population.

Organizing data

Datasets order observations and variables in the form of rows (observations) and columns (variables). This is often referred to as the **N×K design**, where 'N' refers the number of observations and 'K' refers to the number of variables.

In this dataset, for example, Observation 1 (in Figure 3.1, 'obs1') is what we are interested in investigating: for example, a person, a business, a group, or a country. Then the value of the first variable ('Var1') for Observation 1 goes in the row for 'obs1' and column for 'Var1'. Then the value of the second variable ('Var2') for Observation 1 goes in the row for 'obs1' and the column for 'Var2'. And on until we have completed the row of Observation 1. We have input data into the dataset. We can continue to do the same for Observation 2. And repeat until we have run out of observations.

Voilà, a dataset!

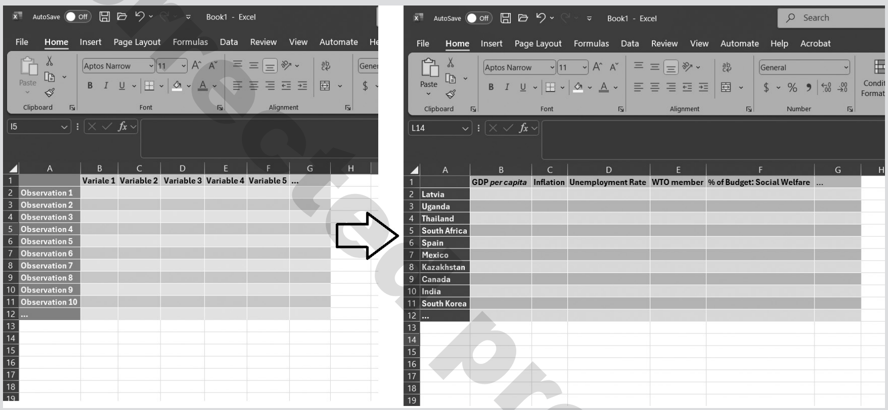

Figure 3.1 A Dataset

Source: Microsoft Excel spreadsheet (for more information: microsoft.com). Used with permission from Microsoft

Two: Data should be what we want to measure

For scientific research and statistical analysis, data must be transformed into meaningful representations of what we intend to study. The data that we intend to use must conceptually and operationally capture the phenomena that we are interested in. That is, data become useful to research by being transformed into variables.

Variables are specific collections of data that represent concepts in which we are interested. Variables will need to represent what we are interested in, that is, they will need to connect the concepts under investigation with 'real-world' empirical referents.

Some concepts offer clear, easy to observe indicators such as the number of votes a party got in an election, GDP *per capita*, or the number of military bases. In this case, compiling and aligning data with the intended variable is often straightforward. However, variables must also represent less concrete concepts we are interested in, such as force, justice, trust, or inequality. Unlike before, researchers may struggle to convert data into the variables necessary to test and control in their analysis. We must also take the final step of locating a reasonable, measurable, and empirical indicator.

These are data

An increasingly available and valuable source of data in research is text. Whether drawn from news articles, social media posts, legislative records, or other documents, text can be transformed into structured data through techniques like word counts, sentiment analysis, and topic modelling. From a statistical perspective, text-as-data variables function like any other variable – whether we are analysing GDP *per capita* or the frequency of economic terms in a body of pronouncements – the same fundamental principles of inference, uncertainty, and hypothesis testing apply. The growing accessibility of digitized text expands opportunities for measuring public opinion, political discourse, and cultural trends with rigorous statistical tools.

At the same time, social scientists are also, in a sense, overrun with variables that must be contended with. The complexity of world, whether the social reality of influences on vote choice for individuals or the international contexts of the emergence of democracy, is subject to several potential influences. At the same time, we must contend with data quality: data should be sufficient to contain the information we need on all relevant variables to answer the research question. Not only is it important to have correctly collected, measured, and codified the phenomena that you are interested in, but also everything else that might impose upon that relationship. This is often the hardest criterion.

Fortunately, there are meaningful solutions to each of these challenges. The complexity of variables explains the attractiveness of statistics, namely, its ability to deal with complexity quite handily (Chapter 4). And despite the contests over definitions and 'real-world' indicators, the process of identifying and measuring the phenomena in which we are interested is fairly straightforward.

Let's start from here.

Conceptualization and operationalization

Let's assume that we want to tackle a difficult and timely issue in Political Science. We decide to study the effect of inequality on democracy. Big bite, lots going on here. But, please note, when we say that we are studying inequality and democracy, we aren't really. At least not very well.

You and I and everyone who does Political Science are not evaluating actual inequality or probing the guts of democracy. We instead are evaluating the best indicators – derived from the available data – of the concepts of inequality and democracy. Instead of putting inequality and democracy into a laboratory, we devise a concept of what we are interested in and explicitly define what we intend to study. Then, we find an indicator that most closely captures this definition.

These two ideas – 'defines' and 'captures' – are possibly the most important steps in the scientific method. **Conceptualization** – the process of defining – and **operationalization** – the process of capturing – constitute the link between our ideas and the real world and are thus crucial elements in the research process. Despite their importance, the processes of conceptualization and operationalization are not perfect and are rarely fixed.

Recall the pilot metaphor in which subjective inputs not only take place in the objective process of the scientific method but also *must* take place as key elements in the quality and safety of the flight. For us, the (subjective) choices made at this stage will determine (1) which analytical techniques are available to your research, (2) the results of your research, and (3) your ability to make (inferential) claims beyond the analysed data of your research.

Sounds important.

It is.

For our study on inequality and democracy, not only has it proven quite difficult for everyone working on inequality to agree on the most appropriate definition of inequality (or democracy), but it has also been difficult to then create (or find) a measure – or 'indicator' – that satisfies the preferred definitions. Yet, research on the relationship between democracy and inequality continues apace. How?

Remember the theory as model discussion? We don't need every bit and bob on the model to distinguish it from similar and dissimilar models. We need what are essential to the identification of that concept.

And then we just need something to measure that. Now, let's try our hand at conceptualizing and operationalizing.

Data are not your concept

How you define (*conceptualize*) and measure (*operationalize*) your variables has a profound effect on what you will find.

– Every statistics textbook ever

Conceptualization

Conceptualization is the process of defining the concepts in which we are interested (or need for control). If you want to be all philosophical about it, conceptualization is the delicate application of a metaphysical knife to the continuous current of reality to categorize what we perceive as its separate parts in order to produce discrete, definable observations. Or if we are keeping it in the streets, conceptualization tells us what your concepts are – and what they are not.

This process has two competing demands that simultaneously stretch and constrain our concept. First is *connotation*. What do you mean by the word you are using? What are the defining characteristics or properties? Connotation challenges us to identify and remedy any ambiguity we might have about what we are studying. We want to minimize the distance between the word and what we intend. This is the stretch of our defining process. We want to be sure that everything we intend to study is included.

The second demand, *denotation*, refers to objects themselves. To what are you referring? Where does our definition meet its limit? We should confront any vagueness between our meaning and what defines the extension of the definition. Denotation constrains our definition to only those things we want to study.

Like theoretical parsimony, the competing demands of connotation and denotation force us to confront what we intend to include and exclude from our concept. This is a refinement process, carefully inspecting our own ideas about what we intend to study and what that includes. And excludes, so that we are not referring to things we didn't intend to include.

Now, recall our potential research on inequality and democracy. Or economic development, poverty, justice, political apathy, partisanship, conflict, party competition, or power. How do we unambiguously move from just words on the page to what we intend to include – and exclude – in our definition? For precisely this reason, conceptualization often takes up lots of space in the design of a research project. Literal space, as articles and books and other publications often spend a great deal of time articulating what it is the research will examine and how we will measure it.

A Political Scientist took this process seriously – motivated by what he perceived was weak conceptualizing – and proposed a rubric for remedying such issues. In 1970, Giovanni Sartori published an article in the *American Political Science Review* in which he highlighted the importance of thinking about our concepts, reminding researchers – as I have pointed out above – that what one measures shapes the outcomes of one's research. He added, what one *claims to measure* does as well, cementing this idea in the minds of its readers with this beauty: 'We are deluding ourselves if we really believe that by *saying* a variable, we *have* a variable' (Sartori 1970: 1037, *emphasis mine*).

He advised researchers to avoid 'conceptual stretching', thereby undermining the results. As a guide, Sartori's suggested using a 'Ladder of Abstraction'. A ladder, in perspective, is wider at the end closest to you and narrower at the far end. With these inverse proportions in mind, he argued that as the attributes assigned to a concept increase (an increase in the number of rungs up the ladder), its extension (the other end) narrows. That is, the more specificity – the more dimensions or characteristics – we include in our concept, the more objects we exclude. While an imperfect guide – but killer metaphor – he was suggesting that researchers attempt to make extensional gains (denotation) without losing the precision and empirical testability (connotation). Or simply, to extend as far as the concept will go, but not farther (Einstein, natch).

In a sense, Sartori's conceptualization is a Goldilocks equilibrium. We may have to try concepts too big and too small before alighting on the one that is just right. What we must avoid, in any case, is a compliment from the great Greek philosopher Mediocrates, 'Eh… good enough.'

Operationalization

Once we have satisfactorily defined our concept, the second step is operationalization. Operationalization is the process of matching our concept to useable, empirical referents. What is the actual evidence? What is it we are going to look at?

Imagine if you had designed a wonderful research project, and even gotten as far as conceptualizing all of the relevant variables. Then, you fell ill and couldn't leave the house (which mysteriously has no internet in this story). You call your friend, who runs to the window outside your house, and you shout down from the window to ask her to help with operationalizing your concepts for your research project. You tell her, 'Bring me X, a list of Ys, and a dataset with Zs.' Operationalization is the process of naming X, Y, and Z. It is the specific, tangible thing that you will actually measure. It is what will be in your friend's hands when she returns.

Operationalization is shaped by two things: the quality of conceptualization and the considerations of the availability, sufficiency, and suitability of data (as potential indicators). Fortunately, the process of conceptualization can do a great deal of the operational work for you. The better the definition, the greater the likelihood of a better choice of indicator.

Developing development

In 1990, the United Nations Development Programme (UNDP) embraced a new type of measurement for 'human development'. While the previous indicator, *Gross National Product* per capita, may have represented a reasonable proxy for several dimensions of human welfare, the UNDP wanted to measure development in a means that captured 'the process of enlarging people's choices' (Sen, 1995). So, the UNDP started including adult literacy and combined enrolment ratios, life expectancy at birth, and an adjusted GNP. They argued that these were essential for people to lead a long and healthy life, to acquire knowledge and have access to resources for a decent standard of living. This change to the conceptualization and operationalization of human development has non-trivial impacts on not only how we *think* about human development – but also real-world consequences of *where and how much assistance is assigned*.

As an aside, I am still assuming you want to impress potential new friends at parties – so just drop 'conceptualization' and 'operationalization' casually into conversation to raise the bar for any competitors. You can borrow this: 'Hey, how you doin'? Wouldn't you agree that the difficulty of operationalization is problematized by weak conceptualization?' (This can also be used to deflect half-soaked, smart-ass uncles at a family dinner who don't think much of you going to university.)

Let's look at an example with the quality of conceptualization and then turn to data considerations.

Example: Conceptualization and operationalization

It is not possible to overstate the consequence of these two steps in our research. Conceptualization and operationalization represent our selection of the variables that will be analysed to produce the results from which we will draw our conclusions and inferences. Errors or laziness at this stage may fatally undermine the goal of our scientific endeavour.

The good news is that most of the time, conceptualization and operationalization of the key variables have already been done for you. The literature to which you are writing very likely has fought for – and settled – on agreed-upon concepts and indicators for what you want to analyse. However, in some cases, one stumbles upon a question or literature that might be shaken up with a slight change of concept or slight change in indicator.

There is no one answer for a problem of conceptualization and operationalization. What is one to do when faced with a conceptualization challenge or one recognizes the possibility of conceptual stretching?

- Argue for a new definition?
- Search for another concept?
- Ignore it?
- Research something else, something more concrete?

Clearing that hurdle, what do we do when searching for an indicator?

- Which indicator do we choose?
- Do we create our own?
- Do we measure something else to include or exclude different characteristics?

These are not moot questions in Political Science. Let's revisit political participation. The typology, on which the edifice of the study of political participation is built, derives from a series of articles in the late 1970s and early 1980s. These contributions shared a generic definition (or *conceptualization*) of political participation as the actions of individuals (either individually or in groups) to affect the policy-making bodies through conventional and unconventional political methods.

Among these, a seminal article articulated the shared conceptual premise of political participation in advanced industrial democracies as a typology of actions. Verba et al.

(1978) argued for a typology of ‘conventional’ political participation – including voting, working in an election campaign, and attending rallies; ‘unconventional’ political participation – including legal boycotts, picketing, strikes, etc.; and ‘illegal acts’ – including rioting, unauthorized demonstrations, etc. Perhaps information about individuals’ participation in illegal acts were harder to elicit in surveys: ‘Next question, would you firebomb a city council building? Yes or no? And please answer honestly.’

This became the operational basis for nearly all subsequent studies of political participation in the West. That is, the vast majority of studies of political participation collected data on individuals’ choices of conventional and unconventional political actions (usually in the form of surveys asking respondents to report their having done something or willingness to do so).

The nature of political participation however is changing. One can choose to make sitting down to dinner a political act by the choice to be a vegetarian in opposition to industrial farming. Not wearing fur, reclaiming urban public space, flash mobs, working at a food shelter, hosting a podcast, participating in political theatre, wearing – or not wearing – a mask during a pandemic are all modern equivalents of political protest (read: participation).

This is what Theocharis and Van Deth (2018) argue. Updating our thinking about political participation, they offer a new taxonomy of political participation that isn’t, ‘Yeah, online counts, too, you know.’ While Theocharis and Van Deth do suggest that the internet is just the extension of old forms in a new platform, what is new, or transformative, in their words, are the creative, expressive, and individualized modes of protest activities. These newer forms share the key conceptual dimensions of conventional and unconventional participation, but appeal to younger, more politically disaffected citizens seeking non-institutional participation on their own terms.

This new taxonomy presents us with challenges to the decades-old definitions and measures – the conceptualization and operationalization – of political participation. Given that our choice of definitions and indicators dictate what we will find, what should we do with political participation? There is no single answer.

Do we need to change the definition? The measures? Both?

Perhaps conceptually, this is not overtly difficult. Can newer forms of political actions be ‘actions of individuals (either individually or in groups) to affect the policy-making bodies through conventional and unconventional political methods’?

A tentative yes.

But then how do we then measure the newer forms of political participation? Or more precisely, how do we include (read: ‘count’) the array of ‘creative, expressive, and individualized modes of protest activities’? This poses a more difficult challenge. How do we know if someone’s vegetarianism, mediocre one-man political comedy show, learning Rage Against the Machine songs on the guitar, homemade sign pointing at an unfixed pothole, delivering food to a poor neighbourhood, not getting vaccinated are political actions intending to ‘influence the outcome of a political process’ and should be assessed? Do they need to be impactful? Intentional? Authentic? Or are these simply the function of an inflated sense of self-importance, misinformation, or misguided thinking – and thus excluded?

That is, operationally, how do we distinguish – and count these – as political participation?

That's a bit harder to say.

This case of political participation presents us with a pretty serious challenge. Good. In fact, this might be a useful riposte to the challenge that Political Scientists seem to be always looking at the same questions. If times are a-changing, we should learn how to adjust our research as well. Here, as in so many aspects of scientific study, there is no one answer, or no uniformly *best* answer. The answer will be up to the researcher(s) who choose to take up this challenge.

Let us then confront two crucial attributes of potential indicators: **validity** and **reliability**.

Lost in translation

Finding the right definition and indicator is made harder as we often want to compare concepts across contexts differentiated by political or economic institutions as well as political cultures and societies. We must consider how universal our concept and measures are for comparison. This is called **conceptual congruence**.

Conceptual congruence takes conceptualization and operationalization on vacation. Is what we want to measure – and how we are measuring it – consistent across all groups (e.g., regions or countries)? How universal is our concept? Can we assume some level of cross-unit conceptual consistency?

A good example of conceptual congruence is (still) political participation. Many studies on political participation use the questions we have seen above. Four of those are signing a petition, attending lawful demonstrations, joining in boycotts, and joining strikes. This seems simple enough. So what's the problem?

Joining strikes is deeply culturally driven and very often distinct. In the UK, some strikes call for 'working to contract' or 'work-to-rule' strikes. This is similar – although only similar – to the Italian 'white strike' ('sciopero bianco'). In both, workers do no more than what they are contractually obliged (although UK workers are often forced to make up this work). It is meant to show how much of their job, the quality of their job, is related to the 'extra' they put in by enjoying the work, their colleagues, or other elements of their workplace. By contrast, this concept – along with striking in general – does not exist or is not widespread in the US (in which much is tied to employment, e.g. access to healthcare). One might argue that striking at all in the US is so uncommon as to attract national attention versus the commonplace British and Italian strikes. So, a Brit, Italian, and American saying 'don't mind if I do/si, dai/hell yeah' to 'joining a strike' represent three potentially different responses to this form of participation *in meaning*.

We must ask ourselves, does this conceptual variation across countries create any problems for our analysis of political participation? Perhaps it could. Is this a problem for our analysis? Perhaps it is.

Validity and reliability

The indicators that are available to us are, as we started this chapter with, the data that exist or can be gathered. The ones that we choose should satisfy two criteria. First, it is important that the indicator you choose makes intuitive sense. If you intend to find an indicator of inequality (or democracy), the indicator you choose should be a reasonable measure of inequality (or democracy). This intuitive sense refers to whether the indicator is *valid* in the Sartorian sense: does it actually measure the concept you say it does?

One approach to establishing operational validity is to argue that it is. That is, one can make a case for why this indicator should be valid 'on its face' (sometimes called, creatively, *face validity*). This is not an easy approach but it is not a wrong approach, especially in the context of a change in concept, something novel, or the discovery of something unexpected. Another approach, perhaps more convincing than the previous, is to provide *construct validity*. If two concepts are related, then it stands to reason that the measures of those concepts should also be related. Demonstrating that the two measures are in fact related underscores the claim that we are measuring what we intend to measure.

Second, our indicator must be *reliable*. Reliability is the expectation that the indicator you choose will give you consistent results. That, measured again and again, the indicator isn't capricious, unstable, or unpredictable.

Establishing indicator reliability can be simpler although for new indicators, it can be a challenge. One approach is to *test/re-test*. That is, in repeated measures of the same phenomenon, we should get more or less the same thing time after time. This is a common approach for measures found in over-time datasets or repeated samples. Another approach is a *split test*. In random subsets of your sample, our indicator should look more or less the same. That is, if we repeatedly measure our indicator on groups of our data, we should get roughly similar results. Both of these approaches are not rules – but rather 'rules of thumb'.

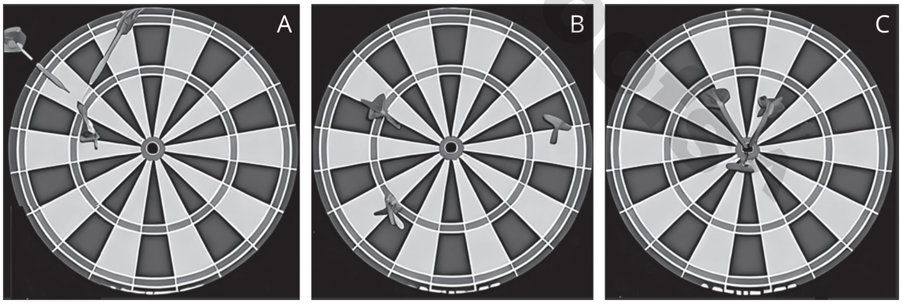

Figure 3.2 Validity and Reliability

While validity and reliability are closely related, they do bring separate pieces of information to the table. Above are illustrative representations of validity and reliability. Let's assume that we are aiming for the bullseye – our intended concept – and the darts represent an indicator or operationalization of that concept. One could substitute accuracy – hitting the correct target – for validity; and precision – returning the same results again and again – for reliability. In Figure A (of Figure 3.2), the darts are reliable but not valid. The darts congregate consistently – reliably – in the same place but miss the intended target (i.e., are invalid). In Figure B, our indicator is neither valid nor reliable as it has missed completely. Finally, in Figure C, our indicator is both reliable and valid. This is the indicator we need.

Example: Validity and reliability

One of the most challenging questions in Political Science is to locate parties along the ideological continuum. One approach has been to simply ask country experts to rank each party in their countries. But, one is left wondering, how good are these expert placements? How valid and how reliable are these expert placements of parties along the ideological continuum?

One approach for reliability, taken from above, is test and re-test. Ask the experts to place the parties along the continuum. And then – later – literally ask them to do it again. Similar results suggest reliability. Alternatively, one could have *more* experts make placements. High similarity among a larger set of experts suggests a high level of *inter-coder – thus indicator – reliability*. Beyond ideological placement, this is a common approach in political communication or text and content analysis.

Other way to establish validity and reliability is to construct an independent (or at least, alternative) means to measure your concept. To the extent that these multiple measures align, it follows that the indicator you choose is likely a valid one. In our example, if we assume that citizens will most likely vote for parties located near their own ideological positions (please note, hotly debated!), one could replace the ideological location of parties with the average ideological location of the citizens that voted for them. So, for example, if extreme left parties' ideological positions align with far-right voters' ideological positions, there may be an issue with the validity of your indicator. If, however, voters' ideological locations align more or less with expert placement of parties' ideological locations, your measures gather greater support for their validity.

Data: An origin story

The Data-Generating Process (DGP) is the true, underlying phenomenon that creates the data, while a model is an often-imperfect attempt to approximate this process. The DGP refers to three key aspects: (1) the actual mechanisms by which data are produced, (2) the statistical approximation of this process, and (3) the probability model for any data creation procedure. For example, in a simple case like rolling dice, many factors influence the outcome – symmetry, starting orientation, force, surface shape, friction, air movement, and more. In reality, we cannot measure every single factor, which is why we rely on probability models to capture the uncertainty in data generation. This means the true DGP remains unobservable, and we can only model it indirectly.

Data quality

Once we have assembled our ideas about what data we might need/want, we are confronted with the trifold problem of data quality in the 'real world' of the researcher: data availability, sufficiency, and suitability.

Data availability

Operationalization is naturally constrained by the availability of data in the material sense that we must select from what is best available to capture our concept. Very often one can use what is common to the literature (i.e., what is agreed upon). While this seems the path of least resistance, there remain the practical constraints of doing so. Existing measures are often used for their convenience and relative abundance (as well as 'acceptable' level of problems). This does not mean it is incorrect, it just means that you'll need to be aware of the limitations of doing so.

As a quick example, in our inequality and democracy research, there are multiple commonly used measures of democracy available:

- The *World Governance Indicators* (https://info.worldbank.org/governance/wgi/) from the World Bank in over 200 countries spanning nearly 25 years. These include the process by which governments are selected, monitored, and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them.
- *Freedom House's* measures of Political Rights and Civil Liberties (now called 'Global Freedom Scores'; https://freedomhouse.org/countries/freedom-world/scores) in more than 200 countries for nearly 50 years. These measure *individual freedom* – the opportunity to act spontaneously outside the control of the government in two broad categories: political rights and civil liberties.
- *The Polity IV Project* (www.systemicpeace.org/polity/polity4.htm), which ranks countries on a scale from –10 (hereditary monarchy) to 10 (fully democratic) for more than 150 countries since 1945. There are six component measures of key qualities of executive recruitment, constraints on executive authority, and political competition; as well as changes in the institutionalized qualities of governing authority.

All of these operationalize the concept of democracy. And they are freely and widely available. Where they differ is the inclusion of the salient attributes.

- What is it about democracy that we *intend*?
- What do we mean when we say 'democracy'?

These measures might move together in a coherent fashion but remain *different* as they represent different *dimensions* of democracy.

By the way, another strategy is to write a grant and get paid to design your own data collection to get exactly the indicators that you want. While this is both a boss move and hugely helpful in getting the data you want, you will be fighting your way back into the established literature and debates (that take place using the more commonly available indicators like above). You will spend a lot of time convincing them that your indicator is in fact the best one. Good luck.

Data sufficiency

Again, unless you are designing the precise process of data collection down to the last question (which is not impossible!), you are at the whim of those that came before you. Ideally, the measures that we choose from the available existing data should reflect the full range of possible values of that concept so that observations are not inadvertently ignored or suppressed. But…

Perhaps you want to study attitudes about the role of women in society. You could use the question from the World Values Survey in which respondents are asked about how 'free' women should be to participate in work and politics. The responses range from 0 'a woman's place is at home' to 12 'women should be equal to men in work and politics' (see, for example, Scarborough et al., 2019). While initially an appealing question, this range of responses limits respondents from being able to register their support for anything *greater* than gender equality. What of those who would prefer to see women take a more prominent role in work and politics than men or support the return of the Amazons of Themyscira? The data are available to study the role of women in society in a large, cross-national survey, yes. But the question is potentially insufficient.

Limited or insufficient data not only constrain the respondent or indicator, but also the comprehensiveness of the analysis. This is another way in which data shapes the process of operationalization in subtle ways.

Data suitability: Units and level of analysis

The **units of analysis** are the subject of your study. We want to correctly identify what it is you intend to study. What is the commonality of what you are studying? Individuals? Parties? Parliaments? Countries? International Organizations?

Recall the hypotheses in which we identified the independent and dependent variables. What are the units of analysis in the following hypotheses?

- More educated citizens are more likely to turn out to vote.
    - *Citizens* are the units of analysis.
- Incumbent MPs are more likely to win re-election than challengers.
    - *Candidates* for parliament are the units of analysis.
- Nations with higher levels of economic development are more democratic.
    - *Nations* are the units of analysis.

At the same time, the ‘data we are seeking’ should match the ‘question we are asking’. The **level of analysis** refers to the stratum or location of the subject of your study. In Political Science, we often think about the two most common: individuals (or micro-level) and nation-states, and supra-national institutions or organizations (macro-level). A third, meso-level, refers to actors in the strata between states and individuals; such as groups or parties.

Maintaining consistency between a theory/hypothesis and the data is a common challenge, particularly when individual attributes and behaviours are ‘explained’ by aggregate patterns. A recognizable example is a television news reporter saying some variation of the following: ‘Regions in which now-mayor Sebastian Laurent won a greater percentage also have larger amounts of downloads of recipes for deep-fried ice cream sandwiches’ – the implication being that those who voted for Sebastian Laurent are the same people downloading recipes for deep-fried ice cream sandwiches.

We – you and me, on the other hand – cannot know whether this is an accurate *aggregate* (i.e., regional) conclusion unless we look at data on *individual* voters which include information both about their vote and online activities. It could very well be that those who voted for Sebastian Laurent are the same people downloading enormous numbers of recipes for deep-fried ice cream sandwiches. Or it could be voters for the other candidates (and abstainers) are disappointed and deciding to just get huge. But, without data at the level of the implication – here, individuals – we cannot determine this.

Counterfactuals are cool, but…

We must also acknowledge when the data are simply unavailable. For example, it might be interesting to know whether UK voters would not have voted for Brexit if they knew that the coronavirus was going to happen. While a potentially fascinating question, there are simply no data that allow us to test such a historical counterfactual. Or, put differently, you would have to build a time machine to find this out. And let’s be clear, if you built a time machine and used it to ask the above question, I only have one thing to say to you. Dinosaurs. Go see the dinosaurs. Or Helen of Troy – most beautiful woman in history, launched a thousand ships? I’ll also accept Jimi Hendrix at the 1970 Isle of Wight festival.

Otherwise, insanity.

Measurement error

The frequency or intensity of wars, the percentage of parliamentary seats, or the number of electorally competitive parties can provide some minor conceptual challenges but can be sufficiently defined and operationalized. Yet, in aligning the concepts under investigation with empirical referents from the real world, there is always some slippage. We looked at democracy and attitudes toward the role of women but would need to account, as two examples, for newer and more accurate forms of gender identification or partisan intensity, for which conceptualization and operationalization present a challenge.

At the same time, we seek to avoid any own goals. We seek to minimize any potential bias in our indicators introduced by measurement bias. While we do not engage data collection methods in this book, operationalization poses one particular problem worth mentioning here. **Systematic measurement error** is often a deviation in our data collection that affects all of the observations in more or less the same way (i.e., systematically). In survey data, it is well known that respondents tend to 'round up' on their income and 'round down' on their age. It would, of course, be good to know the correct number, but again, overwhelmingly the deviations are often relatively small and have few substantial downstream impacts and can be ignored.

In contrast, **non-systematic measurement error** is a problem. The indicator measurements are not slightly biased in one known direction but rather unpredictable. The error in measurement is off, but 'in what direction' and 'by how much' is not an obvious pattern – or is unknown. It doesn't take a lot of imagination to see how measurements that are, well, wrong – and we don't know by how much – are going to produce results that we can't have a lot of confidence in. You are going to want to stop what you are doing and figure out how to get better data.

Systematic error is your dad setting his watch ahead 5 minutes as, in his words, 'it's important to never be late!' While his watch doesn't technically provide the correct time, the correct time can be reliably derived (i.e., by subtracting 5 minutes). Non-systematic error is asking a 3-year-old what time it is. There is no reliable way to determine the correct time from the utterances a 3-year-old gives you (i.e., 'It's Rainbow Dragon time!', '1 fish 2 fish', or 'You have a funny head.').

For the purpose of doing statistics, we might prefer to have an indicator that is reliable but not valid (hopefully a function of systematic measurement error) to an indicator that is valid but not reliable. Why do you think?

A conclusion of sorts

This chapter has confronted the basics elements of the data-generating process and data quality. We focused on a discussion about the measurement of variables including the key issues of conceptualization and operationalization, reliability, and validity. Ultimately, while we should take them seriously, difficulties in conceptualization and operationalization are not necessarily fatal flaws. These choices need to be clearly presented as a means to seek to improve the body of knowledge. In this way, we refine our concepts, their measures, and the resulting analysis.

At some point, researchers have tried their best to identify the key elements of core concepts and draw together what is available to create these indicators. So, first a word of thanks. Second, it *is* a matter of choice. Here, the scientific method converts the researcher into the research itself. Your research will always be a blend of rules to follow and subjective choices. Hence, the emphasis on rigour, control, and transparency. Then others can evaluate how closely you have followed the rules and which choices you have made.

End of chapter summary

- Data are codified observations.
- Data must be representative of what we are trying to study and be what we want to measure.
- Data are not your concept. How you define and measure your variable has a profound effect on what you will find.
- Conceptualization is the process of defining what we want to study.
- Operationalization is the process of finding an appropriate 'real-world' measure for that concept.
- Avoid 'conceptual stretching' and measurement bias.
- The validity of a measure refers to whether we are actually measuring what we say we are measuring.
- The reliability of a measure refers to whether our indicator is a consistent measure of what we want to capture.
- We must strive to maintain consistency between a theory/hypothesis and the data.
- Do you have data? Ask yourself, can it be observed and codified? If yes and yes, then yes.
- Data quality refers to the availability, sufficiency, and suitability of data to satisfy operationalization.
- The units of analysis are the subject of your study.
- We must strive to maintain consistency between a theory/hypothesis and the data.

Glossary

- **Data** are codified observations. **Datum** is singular. **Dataset** is a collection of data.
- **N×K design** is the most common form of a dataset where 'N' is the number of observations and 'K' is the number of variables.
- A **population** is the entire set of what you wish to draw conclusions about.
- A **sample** in a subset of units in the population of interest.
- A **variable** is a collection of data that represent a concept in which we are interested.
- **Conceptualization** is the process of defining the concepts in which we are interested.
- **Operationalization** is the process of measuring our concepts with empirical referents in the real world.
- **Conceptual congruence** is the notion that our concept and its indicator are appropriate for comparison across our units of analysis.
- **Validity** is how well the indicator aligns the concept you intend to measure with the measure you have chosen.

- **Reliability** is the expectation that the indicator you choose for your concept will give you consistent results.
- **Level of analysis** (in Political Science) includes three basic levels: micro- (individuals), macro- (nations or groups of nations), and meso- (whatever falls between such as regions, cities, neighbourhoods, *inter alia*).
- **Units of analysis** are the subject of your study.
- **Systematic measurement error** is a deviation in our data collection that affects all of the observations in more or less the same way (i.e., systematically).
- **Non-systematic measurement errors** are measurements perturbed unpredictably.

Questions

1 Why do we take samples and what kind of problems affect samples?
2 How does the 'competition' between connotation and denotation help us avoid *conceptual stretching* in the process of conceptualization?
3 Name three things – in politics – that seem easy to 'quantify' – that is, they are relatively easy to define and measure. Now, name three that you think would be some of the hardest. Why are they difficult to quantify?
4 We propose a measure for individuals' views on climate change. We argue that the measure is valid but not reliable. Why is this difficult to imagine?
5 One of the major limitations to studying human rights abuse is the difficulty of obtaining data that is not affected by non-systematic measurement error. Why might this be the case?
6 In statistics, would we rather have a measurement that is reliable but not valid (although the difference between the not valid and valid measure is more or less known) or valid but not reliable?
7 Some have argued that the problems associated with conceptualization, operationalization, validity, and reliability in Political Science are often overstated. To what extent should we embrace the notion 'Don't let the perfect be the enemy of the good'?

Annotated references and further reading

Adcock, Robert and David Collier. 2001. 'Measurement validity: A shared standard for qualitative and quantitative research' *American Political Science Review* 95(3): 529–45.

Takes a heads-on approach to remedying some of the most common measurement issues in both qualitative and quantitative research, particularly reaching a consensus and speaking across methodological divides.

Jackman, Simon. 2008. 'Measurement' in Robert E. Goodin (ed.) *The Oxford Handbook of Political Science*. Oxford: Oxford University Press, pp. 119–28.

Jackman presents a more detailed discussion of the issue of measurement. His approach asks a great deal of the reader and spares little in making the case for its importance to the study of politics.

Comparative Manifestos Project: https://manifesto-project.wzb.eu/

For a great example of text and document analysis in European countries.

Bakker, Ryan, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2020. '1999–2019 Chapel Hill Expert Survey Trend File.' Version 1.2. chesdata.eu

For a great example of country experts ranking each party in their countries.

Signposts to research and empirical examples

Collier, David and Steven Levitsky. 1997. 'Democracy with adjectives: Conceptual innovation in comparative research' *World Politics* 49(3): 430–51.

Munck, Gerardo L. and Jay Verkuilen. 2002. 'Conceptualizing and measuring democracy: Evaluating alternative indices' *Comparative Political Studies* 35(1): 5–34.

Ross, Marc Howard and Elizabeth Homer. 1976. 'Galton's problem in cross-national research' *World Politics* 29(1): 1–28. https://doi.org/10.2307/2010045

Sartori, Giovanni. 1970. 'Concept misinformation in comparative politics' *American Political Science Review* 64(4): 1033–53.

Scarborough W.J., R. Sin, and B. Risman. 2019. 'Attitudes and the stalled gender revolution: Egalitarianism, traditionalism, and ambivalence from 1977 through 2016' *Gender and Society* 33(2): 173–200.

Sen, A. (1995). *Human development and economic sustainability* (UNDP Occasional Paper No. 1995/4). New York: United Nations Development Programme.

Theocharis, Y. and J.W. Van Deth. 2018. 'The continuous expansion of citizen participation: A new taxonomy' *European Political Science Review* 10(1): 139–63.

Verba, Sidney, Norman Nie, and Joe-on Kim. 1978. *Participation and Political Equality: A Seven-Nation Comparison*. Cambridge: Cambridge University Press.

Check out the accompanying online resources

Workbooks with chapter exercises and datasets are available online. Want a bit more practice with this chapter's concepts and techniques? Use the companion workbooks online to engage theoretical and practical exercises, available here: https://study.sagepub.com/loveless2e