

# Quantitative Methods and Statistics

true

Version compiled 13 Oct 2020



# Contents

<b>Preface</b>	<b>5</b>
Notation . . . . .	6
License . . . . .	6
Citation . . . . .	6
Technical details . . . . .	6
About the authors . . . . .	7
 <b>Part I: Methodology</b>	 <b>11</b>
<b>1 Introduction</b>	<b>11</b>
1.1 Scientific research . . . . .	11
1.2 Paradigms . . . . .	13
1.3 Instrument validation . . . . .	14
1.4 Descriptive research . . . . .	15
1.5 Experimental research . . . . .	16
1.6 Outline of this textbook . . . . .	19



# Preface

Data are becoming ever more important, in all parts of society, including academia, and including the humanities. The availability of large amounts of digital data (such as text, speech, video, behavioural measurements) raises new research questions, which are typically and often investigated using quantitative methods. Aimed at humanities researchers and students, this book offers an overview of and introduction into the most important quantitative methods and statistical techniques used in the humanities. The book provides a solid methodological foundation for quantitative research, and it introduces the most commonly used statistical techniques to describe data and to test hypotheses. This will also enable the reader to critically evaluate such quantitative research.

This textbook is being used in the course *Methods and Statistics 1* at Utrecht University (Linguistics program). The book is also highly suitable for self-study at a basic level, for everybody who wishes to learn more about quantitative methods and statistics.

The main text has been kept free of mathematical derivations and formulas, which are typically not very helpful for humanities scholars and students. Our explanation is rather conceptual, and rich in examples. Where necessary we present derivations and formulas in separate sections.

This book also contains instructions on how to “do” the statistical analyses and visualisations, both in SPSS (version 22 or later) and in R (version 3.0 or later). These instructions too are in separate sections.

We would like to thank our co-teachers in various courses for the many discussions and examples that have been used in any shape or form in this textbook. We thank our students for their curiosity and for their sharp eyes in spotting errors and inconsistencies in previous versions.

We are also thankful to Gerrit Bloothoof, Margot van den Berg, Willemijn Heeren, Caspar van Lissa, Els Rose, Tobias Quené, Kirsten Schutter and Marijn Struiksmā, for their advice, data, comments and suggestions.

We thank Aleksei Nazarov and Joanna Wall for translating this book from Dutch to English.

Utrecht, October 2020

Hugo Quené, <https://www.hugoquene.nl>

Huub van den Bergh, <https://www.uu.nl/staff/HHvandenBergh>

---

## Notation

Following international usage we use the full stop (decimal point) as decimal separator; hence we write  $\frac{3}{2} = 1.5$ . Note that the decimal separator may vary between computers and between software packages on the same computer. Check which decimal separator is used by (each software package on) your computer.

## License

This document is licensed under the *GNU GPL 3* license (for details see <https://www.gnu.org/licenses/gpl-3.0.en.html>).

## Citation

Please cite this work as follows (in APA style):

Quené, H. & Van den Bergh, H. (2020). *Quantitative Methods and Statistics*. Retrieved 8 Oct 2020 from <https://hugoquene.github.io/QMS-EN/>.

## Technical details

The original Dutch version of this text was written in LaTeX, and was then converted to Rmarkdown, using **pandoc** (MacFarlane, 2020) and the **bookdown** (Xie, 2020) in Rstudio. The Dutch version is available at <https://hugoquene.github.io/KMS-NL/>.

The English translation is based on the Dutch LaTeX version (for Part I) and Rmarkdown version (for Parts II and III).

Other versions of this textbook (EPUB, PDF, HTML), the source code (in Rmarkdown) of the text including examples, accompanying datasets, and figures as separate files, are all available at <https://github.com/hugoquene/QMS-EN/>.

## About the authors

Both authors work at the Faculty of Humanities at Utrecht University, the Netherlands. HQ is professor in the Quantitative Methods of Empirical Research in the Humanities, and he is also founding director of the Centre for Digital Humanities at Utrecht University. HvdB is professor in the Pedagogy and Testing of Language Proficiency, and he is also section chair in Dutch Language and Literature at the Dutch National Board of Tests and Examinations (CvTE).





# Part I: Methodology



# Chapter 1

## Introduction

In this textbook, we will discuss the fundamental concepts, methods, and analytic techniques used in empirical scientific inquiry, both in general and as applied to the broad domain of language and communication. We will look at questions such as: What is a good research question? Which methodology is best for answering a given research question? How can researchers draw meaningful and valid conclusions from (statistical analyses of) their data? In this textbook, we will restrict ourselves to the most important fundamental concepts, and to the most important research methodologies and analytical techniques. In this first chapter, we will provide an overview of various types and forms of scientific research. In the following chapters, we will focus most of our attention on scientific research methodologies in which empirical observations are expressed in terms of numbers (quantitative), which may be analysed using statistical techniques.

### 1.1 Scientific research

To begin, we have to ask a question that refers back to the very first sentence above: what exactly is scientific research? What is the difference between scientific and non-scientific research (e.g., by investigative journalists)? Research conducted by a scholar does not necessarily have to be scientific research. Nor is research by journalists non-scientific by definition just because it is conducted by a journalist. In this textbook, we will follow this definition (Kerlinger and Lee, 2000, p.14):

“Scientific research is systematic, controlled, empirical, amoral, public, and critical investigation of natural phenomena. It is guided by theory and hypotheses about the presumed relations among such phenomena.”

Scientific research is systematic and controlled. Scientific research is designed such that its conclusions may be believed, because these conclusions are well-motivated. A research study can be repeated by others, which will (hopefully) lead to the same results. This demand that research be replicable also means that scientific research is designed and conducted in highly controlled ways (see Chapters ?? and ??). The strongest form of control is found in a scientific experiment: we will therefore devote considerable attention to experimental research (§??). Any possible alternative explanations for the phenomenon studied are looked into one by one and excluded if possible, so that, in the end, we are left with one single explanation (Kerlinger and Lee, 2000). This explanation, then, forms our scientifically motivated conclusion on or theory of the phenomenon studied.

The definition above also states that scientific research is empirical. The conclusion a research draws about a phenomenon must ultimately be based on (systematic and controlled) observations of that phenomenon in reality – for example, on the observed content of a text or the behaviour observed in a test subject. If such observation is absent, then any conclusion drawn from such research cannot be logically connected to reality, which means that it has no scientific value. Confidential data from an unknown source or insights gained from a dream or in a mystical experience are not empirically motivated, and, hence, may not form the basis of a scientific theory.

### 1.1.1 Theory

The goal of all scientific research is to arrive at a theory of a part of reality. This theory can be seen as a coherent and consistent collection of “justified true beliefs” (Morton, 2003). These beliefs as well as the theory they form abstract away from the complex reality of natural phenomena to an abstract mental *construct*, which in its very nature is not directly observable. Examples of similar constructs include: reading ability, intelligence, activation level, intelligibility, active vocabulary size, shoe size, length of commute, introversion, etc.

When building a theory, a researcher not only defines various constructs, but also specifies the *relationships* between these constructs. It is only when the constructs have been defined and the relationships between these constructs have been specified that a researcher can arrive at a systematic explanation of the phenomenon studied. This explanation or theory can, in turn, form the basis of a *prediction* about the phenomenon studied: the number of spoken languages will decrease in the 21st century; texts without overt conjunctions will be more difficult to understand than texts with overt conjunctions; children with a bilingual upbringing will perform no worse at school than monolingual children.

Scientific research comes in many kinds and forms, which may be classified in various ways. In §??, we will discuss a classification based on paradigm: a

researcher's outlook on reality. Research can also be classified according to a continuum between 'purely theoretical' to 'applied'. A third way of classifying research is oriented towards the type of research, for instance, instrument validation (§1.3), descriptive research (§??), and experimental research (§??).

## 1.2 Paradigms

One criterion to distinguish different kinds of research is on the basis of the paradigm used: the researcher's outlook on reality. In this textbook, we have spent almost all of our attention on the empirical-analytical paradigm, because this paradigm has been written about the most and is the most influential. At present, this approach can be seen as 'the' standard approach, against the backdrop of which other paradigms try to distinguish themselves.

Within the *empirical-analytical* paradigm, we distinguish two variants: positivism and critical rationalism. Both schools of thought share the assumption that there exist lawful generalizations that can be 'discovered': phenomena may be described and explained in terms of abstractions (constructs). The difference between the two schools within the empirical-analytical tradition lies in the way generalizations are treated. Positivists claim that it is possible to make statements from factual observations towards a theory. Based on the observations made, we may generalize towards a general principle by means of induction. (All birds I have seen are also perceived by me to be singing, so all birds sing.)

The second school is critical rationalism. Those within this school of thought oppose the inductive statements mentioned above: even if I see masses of birds and they all sing, I still cannot say with certainty that the supposed general principle is true. But, say critical rationalists, we can indeed turn this on its head: we may try to show that the supposed general rule or hypothesis is not true. How would this work? From the general principle, we can derive predictions about specific observations by using deduction. (If all birds sing, then it must be true that all birds in my sample do sing.) If it is not the case that all birds in my sample sing, this means the general principle must be false. This is called the falsification principle, which we will discuss in more detail in ??.

However, critical rationalism, too, has at least two drawbacks. The falsification principle allows us to use observations (empirical facts, research results) to make theoretical statements (regarding specific hypotheses). Strictly speaking, a supposed general principle should be immediately rejected after a single successful instance of falsification (one of the birds in my sample does not sing): if there is a mismatch between theory and observations, then, according to critical rationalists, the theory fails. But to arrive at an observation, a researcher has to make many choices (e.g., how do I draw an appropriate sample, what is a bird, how do I determine whether a bird sings?), which may cast doubt on the validity of the observations. This means that a theory/observation mismatch

could also indicate a problem with the observations themselves (hearing), or with the way the constructs in the theory (birds, singing) are operationalized.

A second drawback is that, in practice, there are very few theories that truly exclude some type of observation. When we observe discrepancies between a theory and observations made, the theory is adjusted such that the new observations still fit within the theory. In this way, theories are very rarely completely rejected.

One alternative paradigm is the critical approach. The *critical paradigm* is distinguished from other paradigms by its emphasis on the role of society; there is no one true reality: our image of reality is not a final one, and it is determined by social factors. Thus, insight into relationships within society, by itself, influences this reality. This means that our concept of science, as formulated in the definitions of research and theory given above, is rejected in the critical paradigm. Critical researchers claim that research processes cannot be seen as separate from the social context in which research is conducted. However, we must add that this latter viewpoint has lately been taken over by more and more researchers, including those that follow other paradigms.

### 1.3 Instrument validation

As stated above, research is a systematized and controlled way of collecting and interpreting empirical data. Researchers strive for insight into natural phenomena and into the way in which (constructs corresponding to) these phenomena are related to one another. One requirement for this is that the researcher be able to actually measure said phenomena, i.e., to express them in terms of an observation (preferable, in the form of a number). Instrument validation research is predominantly concerned with constructing instruments or methods to make phenomena, behaviour, ability, attitudes, etc. measurable. The development of good instruments for measurement is by no means an easy task: they truly have to be crafted by hand, and there are many pitfalls that have to be avoided. The process of making phenomena, behaviour, or constructs measurable is called *operationalization*. For instance, a specific reading test can be seen as an operationalization of the abstract construct of ‘reading ability’.

It is useful to make a distinction between the abstract theoretical construct and the construct as it is used for measurements, which means: a distinction between the concept-as-intended and the concept-as-defined. Naturally, the desired situation is for the concept-as-defined (the test or questionnaire or observation) to maximally approach the concept-as-intended (the theoretical construct). If the theoretical construct is given a good approximation, we speak of an adequate or valid measurement.

When a concept-as-intended is operationalized, the amount of choices to be made is innumerable. For instance, the Dutch government institute that devel-

ops standardized tests for primary and secondary education, the CITO (Centraal instituut voor toetsontwikkeling, or Central Test Development Institute) must develop new reading comprehension tests each year to measure the reading ability exhibited by students taking the centralized final exams for secondary school students (eindexamens). For this purpose, the first step is to choose and possibly edit a text. This text cannot be too challenging for the target audience, but may also not be too easy. Furthermore, the topic of the text may not be too well-known – otherwise, some students’ general background knowledge may interfere with the opinions and standpoints brought forward in the text. At the next step, questions must be developed in such a way that the various parts of the text are all covered. In addition, the questions must be constructed in such a way that the theoretical concept of ‘reading ability’ is adequately operationalized. Finally, exams administered in previous years must also be taken into consideration, because this year’s exam may not differ too much from previous years’ exams.

To sum up, a construct must be correctly operationalized in order to arrive at observations that are not only valid (a good approximation of the abstract construct, see Chapter ??) but also reliable (observations must be more or less identical when measurement is repeated, see Chapter ??). In each research study, the validity and reliability of any instance of measurement are crucial; because of this, we will spend two chapters on just these concepts. However, in instrument validation research, specifically, these concepts are absolutely essential, because this type of research itself is meant to yield valid and reliable instruments that are a good operationalization of the abstract construct-as-intended.

## 1.4 Descriptive research

Descriptive research refers to research predominantly geared towards describing a particular natural phenomenon in reality. This means that the researcher mostly aims for a description of the phenomenon: the current level of ability, the way in which a particular process or discussion proceeds, the way in which Dutch language classes in secondary education take shape, voters’ political preferences immediately before an election, the correlation between the number of hours a student spent on individual study and the final mark they received, etc. In short, the potential topics of descriptive research are also be very diverse.

---

*Example 1.1:* Dingemanse et al. (2013) made or chose recordings of conversations in 10 languages. Within these conversations, they took words used by a listener to seek “open clarification”: little words like *huh* (English), *hè* (Dutch), *ā?* (Siwu). They determined the sound shape and pitch contour of these words using acoustic measurements

and phonetic transcriptions made by experts. One of the conclusions of this descriptive research is that these interjections in the various languages studied are much more alike (in terms of sound shape and pitch contour) than would be expected based on chance.

---

This example illustrates the fact that descriptive research does not stop when the data (sound shapes, pitch contours) have been described. Oftentimes, relationships between the data points gathered are also very interesting (see §1.1). For instance, in opinion polls that investigate voting behaviour in elections, a connection is often made between the voting behaviour polled, on the one side, and age, sex, and level of education, on the other side. In the same way, research in education makes a connection between the number of hours spent studying, on the one side, and performance in educational assessment, on the other side. This type of descriptive research, in which a correlation is found between possible causes and possible effects, is otherwise also referred to as *correlational research*.

The essential difference between descriptive and experimental research lies in the question as to cause and effect. Based on descriptive research, a causal relationship between cause and effect *cannot* be properly established. Descriptive research might show that there is a correlation between a particular type of nutrition and a longer lifespan. Does this mean that this type of nutrition is the cause of a longer lifespan? This is definitely not necessarily the case: it is also possible that this type of food is mainly consumed by people who are relatively highly educated and wealthy, and who live longer because of these other factors<sup>1</sup>. In order to determine whether there is a causal relationship, we must set up and conduct experimental research.

## 1.5 Experimental research

Experimental research is characterized by the researcher's systematically manipulating a particular aspect of the circumstances under which a study is conducted (Shadish et al., 2002). The effect arising from this manipulation now becomes central in the research study. For instance, a researcher suspects that a particular new method of teaching will result in better student performance compared to the current teaching method. The researcher wants to test this hypothesis using experimental research. She or he manipulates the type of teaching: some groups of students are taught according to the novel, experimental teaching method, and other groups of students are taught according to

---

<sup>1</sup>It is even possible that the nutrition habits under study cause people to live *shorter*, but that this negative effect is masked by the stronger positive effects of education and wealth.



the traditional method. The novel teaching method's effect is evaluated by comparing both types of student groups' performance after they have been 'treated' with the old vs. new teaching method.

The advantage of experimental research is that we may usually interpret the research results as the consequence or effect of the experimental manipulation. Because the research systematically controls the study and varies just one aspect of it (in this case, the method of teaching), possible differences between the performance observed in the two categories can only be ascribed to the aspect that has been varied (the method of teaching). Logically speaking, this aspect that was varied is the only thing that could have caused the observed differences. Thus, experimental research is oriented towards evaluating causal relationships.

This reasoning does require that test subjects (or groups of students, as in the example above) are assigned to experimental conditions (in our example, the old or the new method of teaching) at random. This random assignment is the best method to exclude any non-relevant differences between the conditions of treatment. Such an experiment with random assignment of test subjects to conditions is called a *randomized experiment* or *true experiment* (Shadish et al., 2002). To remain with our example: if the researcher had used the old research method only with boys, and the new research method only with girls, then any difference in performance can no longer just be attributed to the manipulated factor (teaching method), but also to a non-manipulated but definitely relevant factor, in this case, the students' sex. Such a possible disruptive factor is called a confound. In Chapter ??, we will discuss how we can neutralize such confounds by random assignment of test subjects (or groups of students) to experimental conditions, combined with other measures.

There also exists experimental research in which a particular aspect (such as teaching method) is indeed systematically varied, but in which test subjects or groups of students are not randomly assigned to the experimental conditions; this is called *quasi-experimental research* (Shadish et al., 2002). In the example above, this term would be applicable if teaching method were investigated using data from groups of students for which it was not the researcher, but their teacher who determined whether the old or new teaching method would be used. In addition, the teacher's enthusiasm or teaching style might be a confound in this quasi-experiment. We will encounter various examples of quasi-experimental research in the remainder of this textbook.

Within the type of experimental research, we can also make a further division: that between laboratory research and field research. In both types of experimental research, some aspect of reality is manipulated. The difference between both types of research lies in the degree to which the researcher is able to keep under control the various confounds present in reality. In laboratory research, the researcher can very precisely determine under which environmental conditions observations are made, which means that the researcher can keep many possible confounds (such as lighting, temperature, ambient noise, etc.) under control. In field research, this is not the case. When 'out in the field', the re-

searcher is not able to keep all (possibly relevant) aspects of reality fully under control.

---

*Example 1.2:* Margot van den Berg and colleagues from the Universities of Utrecht, Ghana and Lomé investigated how multilingual speakers use their languages when they have to name attributes like colour, size, and value in a so-called Director-Matcher task (Van den Berg et al., 2017). In this task, one research participant (the ‘director’) gave clues to another participant (the ‘matcher’) to arrange a set of objects in a particular order. This allowed the researchers to collect many instances of attribute words in a short period of time (“Put the yellow car next to the red car, but above the small sandal”). The interactions were recorded, transcribed, and subsequently investigated for language choice, moment of language switch, and type of grammatical construction. In this type of fieldwork, however, various kinds of non-controlled aspects in the environment may influence the sound recordings and, thus, the data, including “clucking chickens, a neighbour who was repairing his motorbike and had to start it every other second while we were trying to record a conversation, pouring rain on top of the aluminium roof of the building where the interviews took place.” (Margot van den Berg, personal communication)

---

*Example 1.3:* When listening to spoken sentences, we can infer from a test subject’s eye movements how these spoken sentences are processed. In a so-called ‘visual world’ task, listeners are presented with a spoken sentence (e.g., “Bert says that the rabbit has grown”), while they are looking at multiple images on the screen (usually 4 of them, e.g., a sea shell, a peacock, a saw, and a carrot). It turns out that listeners will predominantly be looking at the image associated with the word they are currently mentally processing: when they are processing *rabbit*, they will look at the carrot. A so-called ‘eye tracker’ device allows researchers to determine the position on the screen that a test subject is looking at (through observation of their pupils). In this way, the researcher can therefore observe which word is mentally processed at which time (Koring et al., 2012). Research of this kind is best conducted in a laboratory, where one can control background noise, lighting, and the position of test subjects’ eyes relative to the computer screen.

---

Both laboratory research and field research have advantages and disadvantages. The great advantage of laboratory research is, of course, the degree to which the researcher can keep all kinds of external matters under control. In a laboratory, the experiment is not likely to be disturbed by a starting engine or a downpour. However, this advantage of laboratory research also forms an important disadvantage, namely: the research takes place in a more or less artificial environment. It is not at all clear to what extent results obtained under artificial circumstances will also be true of everyday life outside the laboratory. Because of this, the latter forms a point to the advantage of field researcher: the research is conducted under circumstances that are natural. However, the disadvantage of field research is that many things can happen in the field that may influence the research results, but remain outside of the researcher's control (see example 1.2). The choice between both types of experimental research that a researcher has to make is obviously strongly guided by their research question. Some questions are better suited to being investigated in laboratory situations, while others are better suited to being investigated field situations (as is illustrated by the examples above).

## 1.6 Outline of this textbook

This textbook consists of three parts. Part I (Chapter 1 to 7) covers research methods and explains various terms and concepts that are important in designing and setting up a good scientific research study.

In part II (Chapters 8 to 12), we will cover descriptive statistics, and in part III (Chapters 13 to 17), we will cover the basic methods of inferential statistics. These two parts are designed to work towards three goals.

Firstly, we would like for you to be able to critically evaluate articles and other reports in which statistical methods of processing and testing hypotheses on data have been used. Secondly, we would like for you to have the knowledge and insight necessary for the most important statistical procedures. Thirdly, these parts on statistics are meant to enable you to perform statistical analysis on your own for your own research, for instance, for your internship or final thesis.

These three goals are ordered by importance. We believe that an adequate and critical interpretation of statistical results and the conclusions that may be connected to these is of great importance to all students. For this reason, part I of this textbook devotes considerable attention to the 'philosophy' or methodology behind the statistical techniques and analyses we will discuss later.

We will also give you instructions on how you can perform these statistical analyses yourself in SPSS (a popular software package for statistical analysis) and in R (a slightly more challenging, but also much more powerful and versatile

software package that has been gaining popularity). For students and employees at Utrecht University, both packages are pre-installed in **MyWorkSpace**. SPSS is available at <https://SurfSpot.nl> for a small fee. R is freely available at <https://www.R-project.org>. A brief introduction to R can be found at <https://hugoquene.github.io/emlar2020/>.

# Bibliography

- Dingemanse, M., Torreira, F., and Enfield, N. (2013). Is “huh?” a universal word? conversational infrastructure and the convergent evolution of linguistic items. *PLOS One*, 8(11):e78273.
- Kerlinger, F. N. and Lee, H. B. (2000). *Foundations of Behavioral Research*. Harcourt College Publishers, Fort Worth, 4th edition.
- Koring, L., Mak, P., and Reuland, E. (2012). The time course of argument re-activation revealed: Using the visual world paradigm. *Cognition*, 123(3):361–379.
- MacFarlane, J. (2020). *Pandoc: a universal document converter*.
- Morton, A. (2003). *A Guide through the Theory of Knowledge*. Blackwell, Malden, MA, 3e edition.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth, Belmont, CA.
- Van den Berg, M., Amuzu, E. K., Essizewa, K., Yevudey, E., and Tagba, K. (2017). Crosslinguistic effects in adjectivization strategies in Suriname, Ghana and Togo. In Cutler, C., Vrzić, Z., and Angermeyer, P., editors, *Language Contact in Africa and the African Diaspora in the Americas: in honor of John V. Singler*, pages 343–362. Benjamins, s.l.
- Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.18.