

Statistics in Psychology Using R and SPSS

DIETER RASCH • KLAUS D. KUBINGER • TAKUYA YANAGIDA

 WILEY

Statistics in Psychology
Using R and SPSS

Statistics in Psychology Using R and SPSS

Dieter Rasch

University of Life Sciences, Vienna, Austria

Klaus D. Kubinger • Takuya Yanagida

University of Vienna, Austria



A John Wiley & Sons, Ltd., Publication

This edition first published 2011
© 2011 John Wiley & Sons, Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Statistics in psychology using R and SPSS / Dieter Rasch, Klaus D. Kubinger and Takuya Yanagida.
p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-470-97124-6 (cloth) – ISBN 978-1-119-97964-7 (E-PDF) – ISBN 978-1-119-97963-0 (O-book) – ISBN 978-1-119-95202-2 (E-Pub) – ISBN 978-1-119-95203-9 (Mobi)

1. Psychometrics. 2. SPSS (Computer file) I. Rasch, Dieter. II. Kubinger, Klaus D., 1949- III. Yanagida, Takuya. [DNLM: 1. SPSS (Computer file) 2. Psychometrics—methods. 3. Statistics as Topic—methods. 4. Software. BF 39] BF39.S7863 2011 150.72'7—dc23

2011020660

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

Print ISBN: 978-0-470-97124-6
ePDF ISBN: 978-1-119-97964-7
oBook ISBN: 978-1-119-97963-0
ePub ISBN: 978-1-119-95202-2
Mobi ISBN: 978-1-119-95203-9

Typeset in 10/12pt Times by Aptara Inc., New Delhi, India

Contents

Preface	x
Acknowledgments	xii
Part I INTRODUCTION	1
1 Concept of the Book	3
References	11
2 Measuring in Psychology	12
2.1 Types of Psychological Measurements	13
2.2 Measurement Techniques in Psychological Assessment	13
2.2.1 Psychological Tests	13
2.2.2 Personality Questionnaires	14
2.2.3 Projective Techniques	15
2.2.4 Systematical Behavior Observation	16
2.3 Quality Criteria in Psychometrics	16
2.4 Additional Psychological Measurement Techniques	17
2.4.1 Sociogram	17
2.4.2 Survey Questionnaires	17
2.4.3 Ratings	18
2.4.4 Q-Sort	18
2.4.5 Semantic Differential	19
2.4.6 Method of Pair-Wise Comparison	19
2.4.7 Content Analysis	19
2.5 Statistical Models of Measurement with Psychological Roots	20
References	20
3 Psychology – An Empirical Science	22
3.1 Gain of Insight in Psychology	23
3.2 Steps of Empirical Research	26
References	29

4	Definition – Character, Chance, Experiment, and Survey	30
4.1	Nominal Scale	35
4.2	Ordinal Scale	35
4.3	Interval Scale	37
4.4	Ratio Scale	38
4.5	Characters and Factors	40
	References	41
Part II DESCRIPTIVE STATISTICS		43
5	Numerical and Graphical Data Analysis	45
5.1	Introduction to Data Analysis	45
5.2	Frequencies and Empirical Distributions	49
5.2.1	Nominal-Scaled Characters	50
5.2.2	Ordinal-Scaled Characters	55
5.2.3	Quantitative Characters	63
5.2.4	Principles of Charts	73
5.2.5	Typical Examples of the Use of Tables and Charts	74
5.3	Statistics	77
5.3.1	Mean and Variance	77
5.3.2	Other Measures of Location and Scale	79
5.3.3	Statistics Based on Higher Moments	91
5.4	Frequency Distribution for Several Characters	94
	References	97
Part III INFERENCE STATISTICS FOR ONE CHARACTER		99
6	Probability and Distribution	101
6.1	Relative Frequencies and Probabilities	101
6.2	Random Variable and Theoretical Distributions	107
6.2.1	Binomial Distribution	109
6.2.2	Normal Distribution	116
6.3	Quantiles of Theoretical Distribution Functions	123
6.4	Mean and Variance of Theoretical Distributions	125
6.5	Estimation of Unknown Parameters	126
	References	129
7	Assumptions – Random Sampling and Randomization	130
7.1	Simple Random Sampling in Surveys	132
7.2	Principles of Random Sampling and Randomization	134
7.2.1	Sampling Methods	134
7.2.2	Experimental Designs	140
	References	146
8	One Sample from One Population	147
8.1	Introduction	147

8.2	The Parameter μ of a Character Modeled by a Normally Distributed Random Variable	148
8.2.1	Estimation of the Unknown Parameter μ	148
8.2.2	A Confidence Interval for the Unknown Parameter μ	150
8.2.3	Hypothesis Testing Concerning the Unknown Parameter μ	156
8.2.4	Test of a Hypothesis Regarding the Unknown Parameter μ in the Case of Primarily Mutually Assigned Observations	165
8.3	Planning a Study for Hypothesis Testing with Respect to μ	169
8.4	Sequential Tests for the Unknown Parameter μ	179
8.5	Estimation, Hypothesis Testing, Planning the Study, and Sequential Testing Concerning Other Parameters	183
8.5.1	The Unknown Parameter σ^2	183
8.5.2	The Unknown Parameter p of a Dichotomous Character	184
8.5.3	The Unknown Parameter p of a Dichotomous Character which is the Result of Paired Observations	189
8.5.4	The Unknown Parameter p_j of a Multi-Categorical Character	192
8.5.5	Test of a Hypothesis about the Median of a Quantitative Character	195
8.5.6	Test of a Hypothesis about the Median of a Quantitative Character which is the Result of Paired Observations	196
	References	199
9	Two Samples from Two Populations	200
9.1	Hypothesis Testing, Study Planning, and Sequential Testing Regarding the Unknown Parameters μ_1 and μ_2	201
9.2	Hypothesis Testing, Study Planning, and Sequential Testing for Other Parameters	214
9.2.1	The Unknown Location Parameters for a Rank-Scaled Character	214
9.2.2	The Unknown Parameters σ_1^2 and σ_2^2	218
9.2.3	The Unknown Parameters p_1 and p_2 of a Dichotomous Character	221
9.2.4	The Unknown Parameters p_i of a Multi-Categorical Nominal-Scaled Character	229
9.3	Equivalence Testing	230
	References	233
10	Samples from More than Two Populations	235
10.1	The Various Problem Situations	236
10.2	Selection Procedures	237
10.3	Multiple Comparisons of Means	238
10.4	Analysis of Variance	241
10.4.1	One-Way Analysis of Variance	241
10.4.2	One-Way Analysis of Variance for Ordinal-Scaled Characters	263
10.4.3	Comparing More than Two Populations with Respect to a Nominal-Scaled Character	265
10.4.4	Two-Way Analysis of Variance	266

10.4.5	Two-Way Analysis of Variance for Ordinal-Scaled Characters	289
10.4.6	Bivariate Comparison of Two Nominal-Scaled Factors	289
10.4.7	Three-Way Analysis of Variance	289
	References	299
Part IV DESCRIPTIVE AND INFERENCE STATISTICS FOR TWO CHARACTERS		301
11	Regression and Correlation	303
11.1	Introduction	303
11.2	Regression Model	308
11.3	Correlation Coefficients and Measures of Association	318
11.3.1	Linear Correlation in Quantitative Characters	318
11.3.2	Monotone Relation in Quantitative Characters and Relation between Ordinal-Scaled Characters	325
11.3.3	Relationship between a Quantitative or Ordinal-Scaled Character and a Dichotomous Character	326
11.3.4	Relationship between a Quantitative Character and a Multi-Categorical Character	330
11.3.5	Correlation between Two Nominal-Scaled Characters	335
11.3.6	Nonlinear Relationship in Quantitative Characters	345
11.4	Hypothesis Testing and Planning the Study Concerning Correlation Coefficients	349
11.5	Correlation Analysis in Two Samples	357
	References	360
Part V INFERENCE STATISTICS FOR MORE THAN TWO CHARACTERS		361
12	One Sample from One Population	363
12.1	Association between Three or More Characters	363
12.1.1	Partial Correlation Coefficient	365
12.1.2	Comparison of the Association of One Character with Each of Two Other Characters	371
12.1.3	Multiple Linear Regression	372
12.1.4	Intercorrelations	374
12.1.5	Canonical Correlation Coefficient	377
12.1.6	Log-Linear Models	377
12.2	Hypothesis Testing Concerning a Vector of Means μ	385
12.3	Comparisons of Means and 'Homological' Methods for Matched Observations	388
12.3.1	Hypothesis Testing Concerning Means	388
12.3.2	Hypothesis Testing Concerning the Position of Ordinal-Scaled Characters	398
	References	400

13	Samples from More than One Population	401
13.1	General Linear Model	401
13.2	Analysis of Covariance	403
13.3	Multivariate Analysis of Variance	414
13.4	Discriminant Analysis	427
	References	445

Part VI MODEL GENERATION AND THEORY-GENERATING PROCEDURES 447

14	Model Generation	449
14.1	Theoretical Basics of Model Generation	449
14.1.1	Generalized Linear Model	450
14.1.2	Model with Latent Variables	453
14.2	Methods for Determining the Quality and Excellence of a Model	454
14.2.1	Goodness of Fit Tests	454
14.2.2	Coefficients of Goodness of Fit	458
14.2.3	Cross-Validation	462
14.3	Simulation – Non-Analytical Solutions to Statistical Problems	464
	References	470
15	Theory-Generating Methods	471
15.1	Methods of Descriptive Statistics	471
15.1.1	Cluster Analysis	471
15.1.2	Factor Analysis	482
15.1.3	Path Analysis	492
15.2	Methods of Inferential Statistics	494
15.2.1	Further Analysis Methods for Classifying Research Units	494
15.2.2	Confirmatory Factor Analysis	501
15.2.3	Models of Item Response Theory	506
	References	518

Appendix A: Data Input 520

Appendix B: Tables 529

Appendix C: Symbols and Notation 538

References 542

Index 547

Preface

This textbook contains, on the one hand, everything that is needed for a freshman statistician. On the other hand, it can also be used in advanced courses and in particular it can be used for empirical research work.

Within the Bachelor's curriculum it is only possible to demonstrate the correct use of the most important techniques. For the Master's curriculum, however, a certain understanding of these methods is necessary. For doctoral studies, understanding alone is not enough: a willingness to reflect critically on the statistical methods must be developed.

Since even for doctoral students a repetition of the basics of statistics on an elementary level is often useful, with this book they can be picked up individually where their powers of recollection end – if necessary at the beginning of the Bachelor education. And in contrast, Bachelor's students are often interested in the contents of a Master's curriculum or where the textbook leads. They can get a taste of that now.

Even lecturers will find something new in this textbook; according to our experience, 'statistics for psychologists' is not taught by professional statisticians but by psychologists, mostly by those at the beginning of their academic careers; anecdotes may at least help them didactically. These casual reflections can of course also be academically amusing for students.

Accordingly, the three to four mentioned target groups are guided through the book using distinctive design elements.

All examples given in this textbook refer to psychology as an empirical science. However, the topics covered in psychology are similar to those of (other) social sciences, above all sociology and educational science. So, of course, this textbook suits their framework as well.

The statistical methods that are recommended in this book and which can be used for answering the research questions posed by psychology as a science are often only practicable when using a computer. Therefore we refer to two software packages in this book. The program package **R** is both freely accessible and very efficient; that is why we continuously use **R** here. However, since in psychology the program package **IBM SPSS Statistics** is still preferred for statistical analyses most of the time, it is also illustrated using the examples; here we use version 19.

We try to present statistical knowledge as simply as possible using these program packages, and avoid formulas wherever reasonable. However, we did not completely avoid formulas because we also wish to help those readers interested in the theoretical background. As a

matter of fact, more important than formulas is the procurement of appropriate applications and interpretations of statistical methods. And that is actually the main focus of this book.

We have refrained from citing the exact sources for the practical, everyday methods given, reserving that for methods that are new or uncommon.

With the hope that the reader may easily gather from this textbook all information relevant to his/her individual academic level.

This book includes an accompanying website. Please visit www.wiley.com/go/statisticsinpsychology

Dieter Rasch, Klaus D. Kubinger, and Takuya Yanagida
Rostock and Vienna

Acknowledgments

The authors wish to express their thanks to those who contributed in either the translation or in the programs for this book, especially:

Dr. Albrecht Gebhardt, Alpen Adria University of Klagenfurth, Austria, who gave us access to the **R**-package **OPDOE** and assistance in programming the sequential triangular tests.

In translating the text from German into English we received help from (in alphabetical order):

Maximilian Alexander Hetzel, University of Vienna, Austria

Nina Heuberger, University of Vienna, Austria

Mag. Jürgen Grafeneder, University of Vienna, Austria

Mag. Bernhard Piskernik, University of Vienna, Austria

Sarah Treiber, University of Vienna, Austria

Mag. Alexander Uitz, University of Vienna, Austria

Because the authors and translators are not native English speakers, we are happy that we found help from (in alphabetical order):

Mag. Carrie Kovacs, University of Luxembourg

Mag. Renate Dosanj, University of Vienna, Austria

Sandra Almgren, Kremmling, Colorado, USA

Peter Loetscher, University of Vienna, Austria

We thank, for a lot of editorial work (in alphabetical order):

Bettina Hagenmüller, University of Vienna, Austria

Mag. Bernhard Piskernik, University of Vienna, Austria

We thank **Prof. Dr. Rob Verdooren, Wageningen, The Netherlands** for carefully reading many chapters and giving helpful remarks, **Dr. Maciej Rosolowski** from the **University of Magdeburg, Germany** for the **R**-programs for the principal component tests.

We further thank **IBM SPSS STATISTICS** for providing the most recent version 19 of the **IBM SPSS STATISTICS** program.

Last but not least, we thank **Mag. Joachim Fritz Punter, Medical University of Vienna, Austria** for producing all figures as a reproduction proof, but above all for editing the antecessor book.

Part I

INTRODUCTION

This textbook requires a multi-layered view of ‘statistics in psychology’. Within the Bachelor’s curriculum it is only possible to demonstrate the correct use of the most important techniques. For the Master’s curriculum, however, a certain understanding of these methods is necessary: for the Master’s thesis, where usually a scientific question has to be worked on single-handed but under supervision, the student has to refer to statistical analyses in literature concerning the topic, and if necessary to improve the choice of the method used for analysis. For doctoral studies, understanding alone is not enough; a willingness to reflect critically on the statistical methods must be developed. The statistical methods used in the doctoral thesis, which means the entrance to a scientific career, have to be oriented on state-of-the-art methodological developments; the ability to follow these developments requires profound knowledge as well as the aptitude to evaluate new statistical methods regarding their shortcomings.

Since even for doctoral students a repetition of the basics of statistics on an elementary level is often useful, with this book they can be picked up individually where their powers of recollection end – if necessary at the beginning of the Bachelor’s education. And in contrast, Bachelor’s students are often interested in the contents of a Master’s curriculum or where the textbook leads. They can get a taste of that now.

Finally, even lecturers will find something new in this textbook; according to our experience ‘statistics for psychologists’ is not taught by professional statisticians but by psychologists, mostly by those at the beginning of their academic careers; anecdotes may at least help them didactically. These casual reflections can of course also be academically amusing for students.

Accordingly, the three to four mentioned target groups are guided through the book using distinctive design elements.

The running text, without special accentuation, is directed at all target groups. It is information essential for the further study of the textbook and its practical use – as is this introduction before Chapter 1. Also the terminology used in the book has to be conveyed in a standardized way. Finally, some contents, which should be familiar to doctoral students, are nevertheless aimed at all target groups because we think that repetition is useful.

Moreover, special symbols and labels on the outer edge of some pages signal the target group that the information is aimed at. Target groups other than the ones indicated with the symbol can skip these passages without being in danger of missing the respective educational aim.

The symbol **Bachelor** indicates that the material in these passages is aimed particularly at Bachelor's students since it deals only with the *Ability to Use*. The symbol **Master** on the outer edge indicates that here the reader finds an explanation of the underlying methods, without using a mathematical derivation that is too detailed; this is about *Understanding*. The symbol **Doctor** on the outer edge of the page announces that the shortcomings of the method will be discussed and that common misuses will be indicated; this is about *Critical Reflection*. Finally, the note **For Lecturers** signals didactically useful observations, entailing understanding of the respective topic in a very demonstrative way.

In order to bring all target groups together again, occasionally a *Summary* is given. At the beginning of every chapter a short description of its contents is given.

1

Concept of the book

In this chapter, the structure of the book and accordingly the didactic concept are presented to the reader. Moreover, we outline an example that will be used in several chapters in order to demonstrate the analytical methods described there.

In six sections this book conveys the methods of the scientific discipline of ‘statistics’ that are relevant for studies in psychology:

- I. Introduction (Chapters 1 to 4)
- II. Descriptive statistics (Chapter 5)
- III. Inferential statistics for a single character (Chapters 6 to 10)
- IV. Descriptive and inferential statistics for two characters (Chapter 11)
- V. Inferential statistics for more than two characters (Chapters 12 and 13)
- VI. Theory building statistical procedures (Chapters 14 and 15).

Chapter 1 explains the concept underlying our presentation of the methods. Furthermore an empirical example that will be used as an illustration in various parts of the book is provided.

Chapter 2 will demonstrate that quantifying and measuring in psychology is not only possible but also very useful. In addition we would like to give the reader an understanding of the strategy of gaining knowledge in psychology as a science; the approach however is similar to other scientific fields, which is why this book can be used in other fields too.

In Chapter 3 we will address the issue that empirical research is performed in several steps. For all scientific questions that are supposed to be answered by the study (as diverse as they might be regarding contents), exact planning, careful collecting of data, and adequate analysis are always needed.

Within this context we wish the reader to realize that a study does not always have to include all the people that the research question is directed at. Out of practical reasons, most of the time only part of the group of interest can be examined; this part is usually called sample, whereas the group of interest is called the population. Chance plays an important role here. It will be shown that we have to make probability statements for the results of the statistical analysis; the probability calculus used for this is only valid for events for whose occurrence (or non-occurrence) chance is responsible. For example, a certain event might be that a specific person is part of the study in question. We will treat this topic in Chapter 4, as well as in Chapter 7. Since ‘chance’ often has a different meaning in everyday use as opposed to its general meaning in statistics and therefore in this book, we will point out at this early stage that a random event is not necessarily a rare or unanticipated event.

Finally, if data concerning one or more person(s) or character(s) that are of interest have been gathered within the framework of the study, they have to be processed statistically. The data in their totality are too unmanageable to be able to draw conclusions from them that are relevant for answering the scientific question. Therefore, special methods of data compression are necessary. We will deal with this issue in Chapter 5. The decision of which one of these methods is applicable or most appropriate is substantially based on the type of data: for example, whether they have been derived from physical measurements or whether they can only express *greater/less than* and *equal to* relations. In the latter case it is important to use methods that have been specially developed for this type of data.

Mathematical-statistical concepts are needed, especially for the generalization of study results; these will be introduced in Chapter 6. For readers who are unpracticed in the use of formulas, this chapter is surely difficult, although we try to formulate as simply as possible.

If the generalization of the study results is the aim, then a prerequisite for the use of appropriate methods is that the collected samples are random samples; information on this topic can be found in Chapter 7.

In Chapter 8 an introduction to statistical inference, in particular the principle of hypothesis testing, will be given. Because of the fact that random samples are used, it is necessary to take random deviations of the sample data from the population into account. Through hypotheses that have been formulated before data collection we try to find out as to what extent these deviations are systematic or can/must be traced back to chance. The aim is to either accept or reject a hypothesis based on the empirical data.

Chapter 9 pursues a similar objective, but this time the focus is on two populations that are compared with each other.

The implied separation between planning, data collection, and analysis is true for the classic procedure for empirical studies. In this book, however, we also want to promote a sequential approach. Thereby the gradual collection of data is constantly interrupted by an analysis. This leads to a process that looks like this: observe–analyze–observe–analyze . . . ; this goes on until a predetermined level of precision is reached. This procedure is also described in Chapters 8 and 9.

Special methods are needed in studies that examine a certain character of the research unit (which in psychology often is a person or a group of persons) not only under constant conditions but also under varying conditions or when the study includes more than two populations. In Chapter 10 we cover situations where there are three or more different conditions or two or three treatment factors, with at least two values of each (treatment or factor levels).

In psychological research hardly ever is only one character used. If more than one character per person is observed, then a certain connection between them may exist; we refer here to statistical relationships. If these relationships are of interest, then the statistical methods described primarily in Chapter 11 are needed.

If there really are relationships between several characters – or if there is reason to think so – then one needs very special methods for comparing several populations. Chapters 12 and 13 describe these.

Finally Chapters 14 and 15 give an introduction into theory-building techniques that establish or test models regarding content.

The appendix of the book is split into three parts: Part A lists the data of Example 1.1 which will be illustrated below, and in part B one can find tables, helpful for some analyses; often it is faster and more convenient to look up a value than to calculate it with the help of some software. Appendix C contains a summary of the symbols and abbreviations. A complete list of references and a subject index are given at the end of the book.

Summary

We assume that empirical studies always yield data regarding at least one character. Optimally, planning takes place prior to any study. Data are used to answer a specific question. Statistics as a scientific discipline provides the methods needed for this.

The diverse statistical methods that are recommended in this book and which can be used for answering the research questions posed by psychology as a science are often only practicable when using a computer. Therefore we refer to two software packages in this book. The program package **R** is both freely accessible and very efficient; that is why we continuously use **R** here. However, since in psychology the program package IBM SPSS Statistics is still preferred for statistical analyses most of the time, it will also be illustrated using the examples. The appropriate use of such packages is not trivial; that is why the necessary procedures will be demonstrated by the use of numerical examples. The reader can recalculate everything and practice their use.

The program package **R** can be used for the planning of a study, for the statistical analysis of the data and for graphical presentation. It is an adaptation of the programming language **S** that has been developed since 1976 by *John Chambers* and colleagues in the *Bell Laboratories* (belonging to *Alcatel-Lucent*). The functionality of **R** can be enhanced through freely available packages by everybody and at will, and also special statistical methods and some procedures of **C** and Fortran can be implemented. Packages that already exist are being made available in standardized archives (*repositories*). The most well-known archive to be mentioned here is CRAN (Comprehensive **R** Archive Network), a server network that is serviced by the **R** Development Core Team. With the distribution of **R**, the number of **R** packages has increased exponentially: whereas there were 110 packages available on CRAN in June 2001, there were 2496 in September 2010. **R** is available, free, for Windows, Linux and Apple. With few exceptions, there are implementations for all statistical methods in **R**. With the means of the recently built package OPDOE (see Rasch, Pilz, Verdooren & Gebhardt, 2011), it is possible, for the first time, to statistically plan studies or to calculate the optimal number of examination objects and also to successively collect and analyze data in **R**.

The program package **R** is available for free at <http://cran.r-project.org/> for the operating systems Linux, MacOS X and Windows. The installation under Microsoft Windows is

initiated via the link ‘Windows’, from where the link ‘base’, which leads to the installation website, must be chosen. The setup file can be downloaded under ‘Download R 2.X.X for Windows’ (where X stands for the current version number). After executing this file, one is lead through the installation by a setup assistant. For the uses described in this book all the standard settings can be applied. SPSS as a commercial product must be acquired by purchase; normally universities offer inexpensive licenses for students. More on **R** can be found under www.r-project.org, and on SPSS under www.spss.com. In order not to unnecessarily prolong the explanation of the operational sequence in **R** or SPSS, we always assume that the respective program package, as well as the file that will be used, are already at hand and open.

In **R** the input window opens after starting the program; the prompt is in red: ‘>’. Here commands can be entered and run by pressing the *enter* button. The output is displayed in blue right below the command line. If the command is incomplete, a red ‘+’ will appear in the next line in order to complete the command or to cancel the current command input by pressing the *Esc* button. An instruction sequence is displayed as in the following example:

```
> cbind(sub1_t1.tab, sub1_t1.per, sub1_t1.cum)
```

or also as

```
> cbind(sub1_t1.tab,
+       sub1_t1.per,
+       sub1_t1.cum)
```

or also as

```
> cbind(sub1_t1.tab,
+ sub1_t1.per,
+ sub1_t1.cum)
```

A special working environment in **R** is the *Workspace*. Several (calculation-) objects that have been created in the current session with **R** can be saved in there. These objects include results of calculations (single scores, tables, etc.) and also data sets. A workspace can be loaded with the sequence

```
File - Load Workspace...
```

For all the examples presented in this book the reader can download the *Workspace* ‘RaKuYa.RData’ from the website www.wiley.com/go/statisticsinpsychology.

Since there are more data sets in our *Workspace*, the scores of single research units/persons have to be accessed by specifying the data set with a ‘\$’; for example: `Example_1.1$native_language`. A useful alternative for the access is the

command `attach()`, which makes the desired data set generally available; for example: `attach(Example_1.1)`. To minimize repetition, in the instruction sequences given throughout the book, we assume that the `attach()` command has already been run and therefore the relevant data set is active. For some examples we need special **R** packages; they must be installed once via the menu Packages – Install Package(s)... and then loaded for every session in **R** with the command `library()`. The installation of packages is done via the menu

Packages – Install Package(s)...

In SPSS the desired data frame can be opened via File – Open – Data... after starting the program. Then we write the instruction sequence as in SPSS handbooks; for example like this

Analyze
 Descriptive Statistics
 Frequencies...

For all examples in the book the reader can find the data in the SPSS folder 'RaKuYa' on the website www.wiley.com/go/statisticsinpsychology.

For figures that are shown as the results of the calculations for the examples, we use either the one from SPSS or the one from **R**. Only if the graphs differ between **R** and SPSS will we present both.

It is the concept of this textbook to present illustrative examples with content – that can be recalculated – from almost all subject areas concerning the planning and statistical analysis of psychological studies. A lot of the methods described in this book will be demonstrated using one single data set in order to not have to explain too many psychological problems. This will be introduced in Example 1.1.

Example 1.1 The goal is to test the fairness of a popular natural-language intelligence test battery with reference to children with Turkish native language^{1,2} (see Kubinger, 2009a³).

The following characters were observed per child (see Table 1.1 and the data sheet in Appendix A; then see, for **R**, the respective data structure in Figure 1.1, and for SPSS the screen shot shown in Figure 1.2).

¹Fairness is a specific quality criterion of psychological assessment methods (tests). A psychological test meets the requirement of fairness if the resulting test scores don't lead to a systematic discrimination of specific testees: for example because of sex, ethnic, or socio-cultural affiliation; see Kubinger, 2009b).

²The data originally applied to German-speaking countries; however, there was no socio-political difference when the data in the following analyses were interpreted as relating to English-speaking countries and some ethnic-minority groups.

³Due to copyright reasons the original data had to be slightly modified; therefore no deductions regarding content can be drawn from the data found in the data sheet in the appendix.

Table 1.1 The characters and their names in **R** and SPSS (including coded values).

Name of the character	Name in R	Name in SPSS	Coded values
<i>testee number</i>	no	no	
<i>native language of the child</i>	native_language	native_language	1 = 'German' 2 = 'Turkish'
<i>age of the child</i>	age	age	
<i>sex of the child</i>	sex	sex	1 = 'female' 2 = 'male'
<i>gestational age at birth (in weeks)⁴</i>	age_birth	age_birth	
<i>number of siblings</i>	no_siblings	no_siblings	
<i>sibling position</i>	pos_sibling	pos_sibling	1 = 'first-born' 2 = 'second-born' 3 = 'third-born' 4 = 'fourth-born' 5 = 'fifth-born' 6 = 'sixth-born'
<i>social status (after Kleining & Moore [1968] according to occupation of father/alternatively of the single mother)</i>	social_status	social_status	1 = 'upper classes' 2 = 'middle classes' 3 = 'lower middle class' 4 = 'upper lower class' 5 = 'lower classes' 6 = 'single mother in household'
<i>urban/rural</i>	urban_rural	urban_rural	1 = 'city (over 20 000 inhabitants)' 2 = 'town (5000 to 20000 inhabitants)' 3 = 'rural (up to 5000 inhabitants)'
<i>marital status of the mother</i>	marital_mother	marital_mother	1 = 'never married' 2 = 'married' 3 = 'divorced' 4 = 'widowed'

⁴The gestational age is the age of the (unborn) child counted from the day of supposed fertilization.

Table 1.1 (Continued)

Name of the character	Name in R	Name in SPSS	Coded values
<i>test setting</i>	test_set	test_set	1 = 'German speaking child' 2 = 'Turkish speaking child tested in German at first test date' 3 = 'Turkish speaking child tested in Turkish at first test date'
<i>Everyday Knowledge, 1st test date (T-Scores)⁵</i>	sub1_t1	sub1_t1	
<i>Applied Computing, 1st test date (T-Scores)</i>	sub3_t1	sub3_t1	
<i>Social and Material Sequencing, 1st test date (T-Scores)</i>	sub4_t1	sub4_t1	
<i>Immediately Reproducing – numerical, 1st test date (T-Scores)</i>	sub5_t1	sub5_t1	
<i>Coding and Associating, 1st test date (T-Scores)</i>	sub7_t1	sub7_t1	
<i>Everyday Knowledge, 2nd test date (T-Scores)</i>	sub1_t2	sub1_t2	
<i>Applied Computing, 2nd test date (T-Scores)</i>	sub3_t2	sub3_t2	
<i>Social and Material Sequencing, 2nd test date (T-Scores)</i>	sub4_t2	sub4_t2	
<i>Immediately Reproducing – numerical, 2nd test date (T-Scores)</i>	sub5_t2	sub5_t2	
<i>Coding and Associating, 2nd test date (T-Scores)</i>	sub7_t2	sub7_t2	

⁵ Test scores are generally standardized to a certain scale; T-Scores are a very common method of standardization.

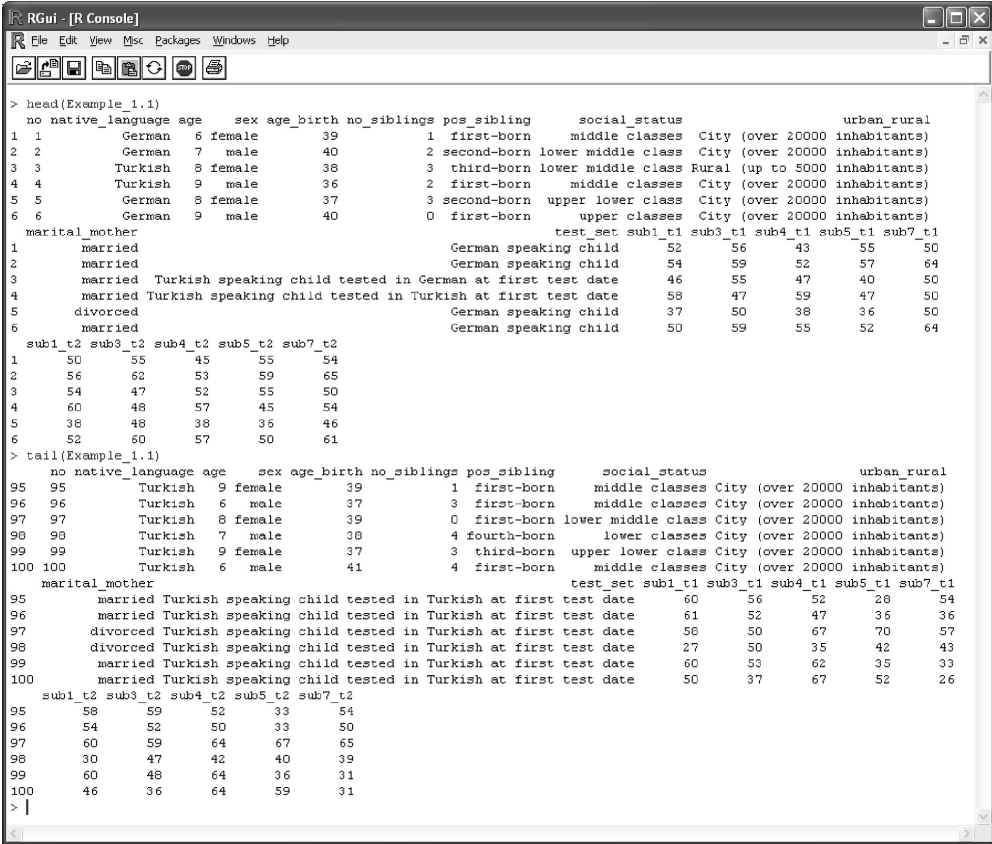


Figure 1.1 Representation of the data structure of Example 1.1 in R.

In order to illustrate some statistical procedures we need other examples regarding content, but the data for these examples will not be found in Appendix A due to space limitations; however they are provided in the aforementioned *Workspace* and *SPSS* folders respectively. For the recalculation of the examples as well as for later calculations with the reader's own data, we will also provide the R instruction sequences, so that they don't have to be typed out. They can be found on the website www.Wiley.com. For beginners in R these are simply listed in order in a PDF file; for those readers already experienced in the use of R they are in a syntax editor for R; that is, *Tinn-R* (www.sciviews.org/Tinn-R/).

	no	native_language	age	sex	age_birth	no_siblings	pos_sibling	social_status	urban_rural	marital_mother
1	1	1	6	1	39	1	1	2	1	2
2	2	1	7	2	40	2	2	3	1	2
3	3	2	8	1	38	3	3	3	3	2
4	4	2	9	2	36	2	1	2	1	2
5	5	1	8	1	37	3	2	4	1	3
6	6	1	9	2	40	0	1	1	1	2
7	7	2	6	1	39	1	1	2	1	2
8	8	2	7	2	36	4	5	3	1	2
9	9	1	9	1	40	0	1	6	1	3
10	10	1	8	2	41	2	2	1	1	3
11	11	1	9	1	36	4	5	2	1	2
12	12	1	6	2	38	1	1	3	1	3
13	13	1	8	1	40	0	1	2	1	1
14	14	1	9	2	41	0	1	3	1	1
15	15	1	6	1	38	1	2	1	2	2
16	16	1	8	2	36	2	1	4	1	3
17	17	1	7	1	40	3	4	2	1	2
18	18	1	9	2	37	2	2	6	1	3
19	19	1	8	1	39	0	1	6	1	3
20	20	1	7	2	35	0	1	4	1	2
21	21	1	6	1	38	4	5	1	1	2
22	22	1	9	2	40	1	1	2	1	3
23	23	1	7	1	41	1	1	3	1	3
24	24	1	6	2	35	0	1	6	2	1
25	25	1	8	1	40	3	4	6	1	3

Figure 1.2 Part of the data view of Example 1.1 in SPSS.

References

- Kleining, G. & Moore, H. (1968). Soziale Selbsteinstufung (SSE): Ein Instrument zur Messung sozialer Schichten [Social Self-esteem (SEE): An Instrument for Measuring the Social Status]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 20, 502–552.
- Kubinger, K. D. (2009a). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch* [Adaptive Intelligence Diagnosticum, AID 2-Turkey Included]. Göttingen: Beltz.
- Kubinger, K. D. (2009b). *Psychologische Diagnostik – Theorie und Praxis psychologischen Diagnostizierens* (2nd edn) [Psychological Assessment – Theory and Practice of Psychological Consulting]. Göttingen: Hogrefe.
- Rasch, D., Pilz, J., Verdooren, R. L., & Gebhardt, A. (2011). *Optimal Experimental Design with R*. Boca Raton: Chapman & Hall/CRC.

2

Measuring in psychology

This chapter deals with several methods of data acquisition that are used in psychology. The methods for psychological assessment and the methods primarily for answering research questions have to be distinguished.

Within the field of psychology, the claim of conducting *measurements*, e.g. to measure ‘psyche’ or psychological phenomena, is often adamantly refuted. The attempt to measure or to quantify would not allow for the specific, individual, and qualitative characteristics of a person. Instead, the assessment of the personality of a person should be performed in a qualitative way.

Psychology as a science demonstrates though that this approach to the assessment of a person, regarding a specific *character* (within psychology: *trait/aptitude*), is limited to a pre-scientific level. While it can lead to important assumptions on causal relations, it never allows for binding generalizations. On the contrary, measurements that are conducted under defined abstractions can relate a person’s personality to an objective framework.

Bachelor Statistical data calls for a useful bundling of what is to be measured. Not everything that is measurable regarding a certain character can be compared in depth, i.e. individually, but the whole essential part of the information has to be compressed. A factually acceptable abstraction of the available information has to be made. For example, this abstraction could be that all 35-year-old women are viewed equally regarding their age, irrespective of whether one of them has a biologically ‘young’ body caused by practicing competitive sports or another one has a biologically ‘old’ body because she lived in war zones for some years.

We can be sure that *measuring* in psychology is valuable for psychological case consulting as well as for research on the evaluation of psychological treatments, and especially for basic psychological research.

Master Doctor

Although there are measurement techniques in psychology that follow the methods of natural science, measurements of psychic or mental phenomena are additionally based on specific scientific methods. One thing, however, is common to all natural sciences, psychology included: measuring means the ascertainment of the interesting character's value for the research unit (in psychology this is mostly a person). This happens as an assignment of numbers or signs in such a way that these assignments (measuring values), represent empirical factual relations. That is, the assignment relations must coincide with the empirical (obviously) given relationships of the research units (discussed in detail in Chapter 4).

Master

Although important to note here but not explicitly a distinct measurement technique are the measurements of physiological psychology: its first sub-specialty, neuropsychology, studies the relationship between behavior and the activity of the central nervous system by the means of electrophysiological methods (e.g. EEG, electroencephalography). Its second sub-specialty, psycho-physiology, investigates the relationships between behavior and the activity of the vegetative nervous system by the means of physical methods (e.g. measurement of electrodermal activity, EDA). Its third sub-specialty, chemical psychology, explores the relations between behavior and chemical substances, which are either brought into the organism from outside (pharmaco-psychology) or are built inside the organism (endocrine psychology, neuro-chemopsychology, psycho-genetics) by the means of chemical methods.

2.1 Types of psychological measurements

Some measurement techniques used in psychology are standard methods of *psychological assessment*; they are used in case consulting of clients but also in research. For these techniques specific psychometric quality criteria apply. Other measurement techniques are used specifically in research. Some of them are also used in fields other than psychology, like sociology or market- and opinion research.

2.2 Measurement techniques in psychological assessment

An extensive introduction to psychological assessment can be found in Anastasi & Urbina (1997) and Kubinger (2009b).

2.2.1 Psychological tests

The term 'psychological test' subsumes all achievement tests, including intelligence tests, as well as so-called objective personality tests.¹

¹If the word 'test' is used without an adjective, the reader should be able to tell from the context if a psychological or statistical test is meant; however, we prefer the use of the more comprehensive term 'psychological assessment tool' when the current explanations aren't limited to (psychological) tests.

Master **Example 2.1** Intelligence test

An example for an intelligence test is the subtest *Verbal Abstraction* from the intelligence test battery AID 2 (*Adaptive Intelligence Diagnosticum, Version 2.2*; Kubinger 2009a). ‘How are a candle and a torch alike?’ or ‘How are an airplane and a bird alike?’ are 2 (very) simple items out of 15 that are presented to the children. The measurement problem is whether the number of solved items alone is a representative and fair measure allowing all the testees to be put into a fair relationship regarding their ability. For example, is child *A*, who solves the first of the aforementioned items but not the very difficult one ‘hunting and fishing’, really as ‘intelligent’ as child *B*, who doesn’t solve the first of the aforementioned items but the last, very difficult one? Specific mathematical statistical methods, developed by *psychometrics*, can answer this question (see, as an introduction, Kubinger 2009b, and more precisely Kubinger, 1989). They can also solve even greater measurement problems in testing. In the AID 2, for example, not all children are given the same items, but items meeting their ability, as demonstrated in preceding items, are selected (so-called adaptive testing).

Master **Example 2.2** Objective personality tests

Objective personality tests register individual stylistic features (‘cognitive styles’) while performing a (achievement) task. The *Gestalt test* (Hergovich & Hörndler, 1994), for example, differentiates between ‘field dependent’ and ‘field independent’ persons. A field-dependent perceptive style occurs when the perceptive environment (field) has a strong influence on the perception target; whereas a field-independent perceptive style occurs when the perception is centered on the perception target. The measurement problems in this test are similar to the aforementioned example: can the number of solved items, in which a figure disguised in confusing line drawings has to be found, put all the testees in a fair relation regarding their field in/dependency?

2.2.2 Personality questionnaires

Personality questionnaires include not only the well-known questionnaires, often called tests, where one rates oneself and one’s typical behavior patterns, but also tools of assessment by others where a third party rates a person as well as ‘tests’ of interest.

Master **Example 2.3** Personality questionnaires

The internationally most renowned personality questionnaire is the MMPI-2 (*Minnesota Multiphasic Personality Inventory-2*; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). ‘I find it hard to keep my mind on a task or job’ and ‘I have had periods of days, weeks, or months when I couldn’t take care of things because I couldn’t “get going”’, are two out of more than 500 statements that have to be rated ‘true’ or ‘false’. The sum of the ‘true’ answers is supposed to indicate the degree of psychasthenia (i.e. a psychological disorder associated with fear and obsessive ideas). We will see below that, in contrast to tests, the measurement problem here is not only aggravated but ultimately completely unsolvable.

Master

Example 2.4 Tools of assessment by others

An often-used tool of assessment by others is the HAMD (*Hamilton-Depression-Rating-Scale*; Hamilton, 1980). Again the problem remains: whether a collection of questions or the scores resulting from them can really picture the empirically given relations. For example, is patient *Z*, who has been rated as a person with suicidal thoughts ('Wishes he were dead or any thoughts of possible death to self') but without any sleeping problems by the psychiatrist or clinical psychologist in charge, as depressive as patient *W* without suicidal thoughts but with occasional '... difficulty falling asleep – i.e. more than $\frac{1}{2}$ hour' and '... being restless and disturbed during the night'? The calculation procedure prescribed by the method, however, presumes this. Of course this presumption would be testable with the above-mentioned methods of psychometrics, but it has not been done yet.

Master

Example 2.5 'Tests' of interest

The above-stated measurement problems of tools of assessment by others also hold for 'tests' of interest, or, even worse, unsolvability becomes a problem, as described above for the questionnaires.

2.2.3 Projective techniques

Projective techniques are psychological assessment tools that try to uncover the personality structure and the motives for action of a person by ambiguous material or stimuli.

Master

Example 2.6 Projective techniques

Projective techniques imply that a person, when confronted with ambiguous material or stimuli, unconsciously reacts with his or her own feelings, thoughts, and attitudes and transfers them into the material. However, the psychometric groundings of projective techniques often do not justify these conclusions. The *Rorschach inkblot test* (see Exner, 2002) is the most well known. As long as the observations of how and as what the testee interprets the (symmetric) inkblots are only of a qualitative nature, the procedure doesn't have a measurement function: users claim that with some disorders it is only through this technique that there is the possibility to be able to talk to the patient, because they open up more easily than if they were asked direct questions. Of course the technique is useful (for the experienced user) in order to hypothesize about behavior-determining conditions or typical facets of a human. However, the technique becomes problematic when the very controversial quantification rules are used. Here the argument mentioned earlier that the numerical relations don't depict the empirical ones doesn't even have to be used, because the technique doesn't even claim to reliably measure any specific character. In order to draw consulting conclusions we have to ask ourselves for example the following question. In which way is person *U*, with 5 holistic answers (that means that the answer can be traced to the registration/description of the whole figure), 3 answers concerning details, and 5 answers concerning small details, better/inferior, stronger/weaker etc. than person *V* with 8 holistic answers, 6 answers concerning details, and 12 answers concerning small details?

2.2.4 Systematical behavior observation

Systematical behavior observation focuses on category-system-bound observations of a person's behavior; it is not the arbitrary or casual observation that is mainly based on subjective impressions.

Master Example 2.7 Systematical behavior observation

Similar to projective techniques, the rule is: as long as only a (qualitative) impression is drawn from a systematical behavior observation that leads to hypotheses on behavior-determining conditions, then the technique has no measurement function. For example in the AID 2 there is a supplemental sheet for the observation of work habits, which is used for the qualitative evaluation of work and contact behavior. Although there are no generally binding quantitative category systems, their use is not connected with measurement problems (for example for the characterization of verbal behavior in social contacts or as a method for the analysis of self-management in everyday life): the counting of observed character categories is similar to physical measurements – for example how often somebody uses the word 'please'.

2.3 Quality criteria in psychometrics

As mentioned earlier, specific quality criteria are relevant for measurement methods in psychological assessment; they define the quality of the collected data: *objectivity*, *reliability*, *validity*, *standardization*, and *unfakeability*.

Objectivity means that the result of the measurement is independent from the diagnosing psychologist. *Reliability* alludes to the degree of formal accuracy of the measurement, i.e. precision. *Validity* refers to the correctness of the measurement with regards to the content, which means that the character that is desired to be measured is actually the one that is measured. *Standardization* permits the placement of the individual measurement result of a person within the distribution of all results of a population. *Unfakeability* means that a measurement instrument doesn't allow individual control of type and content of the desired information. More on psychometric quality criteria can be found in Kubinger (2009b), for example.

Tests usually meet the requirement of objectivity, especially when they are administered in a group setting or via a computer. The exactness of the measurement differs from test to test. In general, tests measure less exactly than physical measurement instruments. Sometimes their validity has not been analyzed, sometimes it is unsatisfactory, but sometimes it is also provided. A (up-to-date) standardization is generally available, and tests are essentially unfakeable.

Personality questionnaires, however, are extremely fakeable: they are very transparent concerning their measurement intentions, and most people will be inclined to answer to their advantage (e.g. socially desired). Consequently personality questionnaires are rarely valid and hardly ever accurate. In a group or computer setting they can be considered objective. This, just as the usually given standardization, doesn't make these kinds of measurement instruments any more useful.

Projective techniques are considerably less fakeable than personality questionnaires because their measurement intention is less transparent. However there are hardly any