

SpringerBriefs in Statistics

For further volumes:
<http://www.springer.com/series/8921>

Ton J. Cleophas · Aeilko H. Zwinderman

SPSS for Starters, Part 2

Ton J. Cleophas
Weresteijn 17
3363 BK Sliedrecht
The Netherlands

Aeilko H. Zwinderman
Rijnsburgerweg 54
2333 AC Leiden
The Netherlands

Additional material to this book can be downloaded from <http://extras.springer.com/>

ISSN 2191-544X ISSN 2191-5458 (electronic)
ISBN 978-94-007-4803-3 ISBN 978-94-007-4804-0 (eBook)
DOI 10.1007/978-94-007-4804-0
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012939200

© The Author(s) 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The small book “SPSS for Starters” issued in 2010 presented 20 chapters of cookbook like step by step data-analyses of clinical research, and was written to help clinical investigators and medical students analyze their data without the help of a statistician. The book served its purpose well enough, since 13,000 electronic reprints were being ordered within 9 months of edition.

The above book reviewed, e.g., methods for

1. continuous data, like t-tests, non-parametric tests, analysis of variance,
2. binary data, like crosstabs, McNemar tests, odds ratio tests,
3. regression data,
4. trend testing,
5. clustered data,
6. diagnostic test validation.

The current book is a logical continuation and adds further methods fundamental to clinical data analysis.

It contains, e.g., methods for

1. multistage analyses,
2. multivariate analyses,
3. missing data,
4. imperfect and distribution free data,
5. comparing validities of different diagnostic tests,
6. more complex regression models.

Although a wealth of computationally intensive statistical methods is currently available, the authors have taken special care to stick to relatively simple methods, because they often provide the best power and fewest type I errors, and are adequate to answer most clinical research questions.

It is time for clinicians not to get nervous anymore with statistics, and not to leave their data anymore to statisticians running them through SAS or SPSS to see if significances can be found. This is called data dredging. Statistics can do more for you than produce a host of irrelevant p-values. It is a discipline at the interface

of biology and maths: maths is used to answer sound biological hypotheses. We do hope that “SPSS for Starters 1 and 2” will benefit this process.

Two other issues from the same authors entitled “Statistical Analysis of Clinical Data on a Pocket Calculator 1 and 2” are rather complementary to the above books, and provide a more basic approach and better understanding of the arithmetic.

Lyon, France, January 2012

Ton J. Cleophas
Aeilko H. Zwinderman

Contents

1	Introduction	1
2	Multistage Regression (35 Patients)	3
	Example	3
	Path Statistics	5
	Two Stage Least Square Method	6
	Conclusion	6
3	Multivariate Analysis Using Path Statistics (35 Patients)	7
	Example	7
	Conclusion	11
4	Multivariate Analysis of Variance (35 and 30 Patients)	13
	First Example	13
	Second Example	17
	Conclusion	19
5	Categorical Data (60 patients)	21
	Example	21
	Conclusion	24
6	Multinomial Logistic Regression (55 Patients)	25
	Example	25
	Conclusion	28
7	Missing Data Imputation (35 Patients)	29
	Example	29
	Regression Imputation	30
	Multiple Imputations	32
	Conclusion	33

8	Comparing the Performance of Diagnostic Tests (650 and 588 Patients)	35
	Example	35
	Conclusion	37
9	Meta-Regression (20 and 9 Studies)	39
	Example 1.	39
	Example 2.	41
	Conclusion	42
10	Poisson Regression (50 and 52 Patients)	43
	Example 1.	43
	Example 2.	46
	Conclusion	48
11	Confounding (40 patients)	49
	Example	49
	Conclusion	53
12	Interaction, Random Effect Analysis of Variance (40 Patients) . . .	55
	Example	56
	Conclusion	60
13	Log Rank Testing (60 Patients)	61
	Example	61
	Log Rank Test.	63
	Conclusion	63
14	Segmented Cox Regression (60 Patients)	65
	Example	65
	Conclusion	68
15	Curvilinear Estimation (20 Patients)	69
	Example	69
	Conclusion	73
16	Loess and Spline Modeling (90 Patients)	75
	Example	75
	Spline Modeling	78
	Loess (Locally Weighted Scatter Plot Smoothing) Modeling.	79
	Note	80
	Conclusions.	80

- 17 Assessing Seasonality (24 Averages)** 81
 - Example 82
 - Conclusions. 85

- 18 Monte Carlo Tests and Bootstraps for Analysis of Complex Data (10, 20, 139, and 55 Patients)** 87
 - Paired Continuous Data 87
 - Unpaired Continuous Data 88
 - Paired Binary Data. 88
 - Unpaired Binary Data. 90
 - Conclusion 90

- 19 Artificial Intelligence (90 Patients)** 91
 - Example 91
 - Conclusion 94

- 20 Robust Testing (33 Patients)** 95
 - Example 95
 - Robust Testing 97
 - Conclusion 98

- Final Remarks** 99

- Index** 101

Chapter 1

Introduction

The first part of this title contained all statistical tests that are relevant for starters on SPSS, and included standard parametric and non-parametric tests for continuous and binary variables, regression methods, trend tests, and reliability and validity assessments of diagnostic tests. Many more statistical methods can be carried out with the help of SPSS statistical software, and the current small e book reviews the most important of them.

We will start with multistep methods for better estimation of multistep relationships (Chap. 2) and multivariate models for assessing data files with multiple outcome variables (Chaps. 3 and 4). Exposure and outcome variables with a categorical rather than linear pattern (Chaps. 5 and 6), and the assessment of missing data (Chap. 7) are the next subjects. Chapter 8 compares the performance of diagnostic tests. Meta-regression and Poisson regression will be reviewed in the Chaps. 9 and 10. Dealing with confounding and interaction using SPSS is in the Chaps. 11 and 12. Survival analyses using Cox methods were reviewed in the first part of this title (Chaps. 15 and 16). This e book reviews log rank tests (Chap. 13) and segmented time-dependent Cox regression (Chap. 14). Various methods for assessing non linear models are in the Chaps. 15 and 16, while autocorrelations for the assessment of seasonality is addressed in Chap. 17. Finally, distribution free methods and robust tests are reviewed in the final three chapters.

Each method of testing is explained

1. using a data example from clinical practice,
2. including every step in SPSS (we used SPSS 18.0 available in many western hospitals and clinical research facilities),
3. and including the main tables of results with an accompanying text with interpretations of the results and hints convenient for data reporting, i.e., scientific clinical articles and poster presentations.

In order to facilitate the use of this cookbook the datafiles of the examples given is made available by the editor on the web through extras.springer.com.

For investigators who wish to perform their own data analyses from the very start the book can be used as a step-by-step guideline with help of the data-examples from the book. They can enter their own data or enter entire data files, e.g., from Excel, simply by opening an Excel file in SPSS, or by the commands “cut and paste” just like with Windows’ Word Program, that everybody knows.

This e book just like part one of this title will be used by the masters’ and doctorate classes of the European College of Pharmaceutical Medicine Lyon France (EC Socrates Project since 1999) as a base for their practical education in statistics, and will be offered together with a theoretical module entitled “Statistics applied to clinical trials”. SPSS statistical software is a user-friendly statistical software with many help and tutor pages. However, we as authors believe that for the novices on SPSS an even more basic approach is welcome. The book is meant for this very purpose, and can be used without the help of a teacher, but the authors are willing to be of assistance for those in search for further help.

De authors are well-aware that this cookbook contains a minimal amount of text and maximal technical details, but we believe that this property will not refrain students from mastering the SPSS software systematics, and that, instead, it will even be a help to that aim. Yet, we recommend that, like with the students in Lyon, it will be used together with the textbook “Statistics Applied to Clinical Trials” by Cleophas and Zwinderman (5th edition, 2012, Springer Dordrecht).

Finally, two last and important points.

1. A data file has rows and columns (vertical rows): the columns are the patient characteristics, otherwise called the variables, 1 row is 1 patient.
2. SPSS software uses commas instead of dots to indicate digits smaller than 1.000.

Chapter 2

Multistage Regression (35 Patients)

Primary question: Is multistage regression better for analyzing outcome studies with multiple predictors than multiple linear regression.

Example

The effects of counseling and non-compliance (pills not used) on the efficacy of a novel laxative drug is studied in 35 patients. The data file is given below.

	Var 1	Var 2	Var 3 (var = variable)
	Outcome	Pt instrumental variable	Problematic predictor
	Frequency counseling	Pills not used	Efficacy estimator of new laxative (stools/month)
1.	24	8	25
2.	30	13	30
3.	25	15	25
4.	35	14	31
5.	39	9	36
6.	30	10	33
7.	27	8	22
8.	14	5	18
9.	39	13	14
10.	42	15	30
11.	41	11	36
12.	38	11	30
13.	39	12	27
14.	37	10	38
15.	47	15	40

(continued)