

Ton J. Cleophas  
Aeilko H. Zwinderman

# SPSS for Starters

 Springer

# SPSS for Starters



Ton J. Cleophas • Aeilko H. Zwinderman

# SPSS for Starters

 Springer

Ton J. Cleophas  
Internist / Clinical Pharmacologist  
Department of Medicine  
Albert Schweitzer Hospital  
Dordrecht, Netherlands  
and  
European College of Pharmaceutical  
Medicine  
Lyon, France  
ajm.cleophas@wxs.nl

Aeilko H. Zwinderman  
Statistician, Department of Biostatistics  
Academic Medical Center  
Amsterdam, Netherlands  
and  
European College of Pharmaceutical  
Medicine  
Lyon, France  
a.h.zwinderman@amc.uva.nl

ISBN 978-90-481-9518-3 e-ISBN 978-90-481-9519-0  
DOI 10.1007/978-90-481-9519-0  
Springer Dordrecht New York Heidelberg London

Library of Congress Control Number: 2010935492

© Springer Science+Business Media B.V. 2010

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

This small book addresses different kinds of datafiles, as commonly encountered in clinical research, and their data-analysis on SPSS Software. Some 15 years ago serious statistical analyses were conducted by specialist statisticians using main-frame computers. Nowadays, there is ready access to statistical computing using personal computers or laptops, and this practice has changed boundaries between basic statistical methods that can be conveniently carried out on a pocket calculator and more advanced statistical methods that can only be executed on a computer. Clinical researchers currently perform basic statistics without professional help from a statistician, including t-tests and chi-square tests. With help of user-friendly software the step from such basic tests to more complex tests has become smaller, and more easy to take.

It is our experience as masters' and doctorate class teachers of the European College of Pharmaceutical Medicine (EC Socrates Project Lyon France) that students are eager to master adequate command of statistical software for that purpose. However, doing so, albeit easy, still takes 20–50 steps from logging in to the final result, and all of these steps have to be learned in order for the procedures to be successful.

The current book has been made intentionally small, avoiding theoretical discussions and highlighting technical details. This means that this book is unable to explain how certain steps were made and why certain conclusions were drawn. For that purpose additional study is required, and we recommend that the textbook “Statistics Applied to Clinical Trials”, Springer 2009, Dordrecht Netherlands, by the same authors, be used for that purpose, because the current text is much complementary to the text of the textbook.

We have to emphasize that automated data-analysis carries a major risk of fallacies. Computers cannot think, and can only execute commands as given. As an example, regression analysis usually applies independent and dependent variables, often interpreted as causal factors and outcome factors. For example, gender or age may determine type of operation or type of surgeon. The type of surgeon does not determine the age and gender. Yet a software program does not have difficulty to use nonsense determinants, and the investigator in charge of the analysis has to decide what is caused by what, because a computer cannot do things like that

although they are essential to the analysis. The same is basically true with any statistical tests assessing the effects of causal factors on health outcomes.

At the completion of each test as described in this book, a brief clinical interpretation of the main results is given in order to compensate for the abundance of technical information. The actual calculations made by the software are not always required for understanding the test, but some understanding may be helpful and can also be found in the above textbook. We hope that the current book is small enough for those not fond on statistics but fond on statistically proven hard data in order to *start on SPSS*, a software program with an excellent state of the art for clinical data analysis. Moreover, it is very satisfying to prove from your own data that your own prior hypothesis was true, and it is even more satisfying if you are able to produce the very proof yourself.

Lyon  
December 2009

Ton J. Cleophas  
Aeilko H. Zwinderman

# Contents

<b>1 Introduction</b> .....	1
<b>2 One-Sample Continuous and Binary Data (t-Test, z-Test) (10 and 55 Patients)</b> .....	3
One Sample of Continuous Data.....	3
Analysis: One-Sample t-Test .....	3
One Sample of Binary Data .....	4
Analysis: z-Test.....	4
<b>3 Paired Continuous Data (Paired-t, Wilcoxon) (10 Patients)</b> .....	7
Analysis: Paired t-Test .....	8
Alternative Analysis: Wilcoxon .....	9
<b>4 Unpaired Continuous Data (Unpaired t-Tests, Mann–Whitney) (20 Patients)</b> .....	11
Analysis: Unpaired t-Test .....	11
Alternative Analysis: Mann–Whitney Test.....	13
<b>5 Linear Regression (20 Patients)</b> .....	15
Simple Linear Regression.....	16
Multiple Linear Regression.....	18
<b>6 Repeated Measures ANOVA, Friedman (10 Patients)</b> .....	21
Repeated Measurements ANOVA.....	21
Alternative: Friedman Test (Better with Non-Gaussian Distributions) .....	23
<b>7 Mixed Models (20 Patients)</b> .....	25
Mixed Effects Analysis .....	27



<b>8</b>	<b>One-Way-ANOVA, Kruskal–Wallis (30 Patients)</b> .....	29
	One-Way-ANOVA .....	30
	Alternative Test: Kruskal–Wallis Test (Better with Non-Gaussian Distributions) .....	30
<b>9</b>	<b>Trend Test for Continuous Data (30 Patients)</b> .....	33
	Trend Analysis for Continuous Data.....	34
<b>10</b>	<b>Unpaired Binary Data (Chi-Square, Crosstabs) (55 Patients)</b> .....	35
	Analysis: Chi-square Test .....	37
<b>11</b>	<b>Logistic Regression (55 Patients)</b> .....	39
	Simple Logistic Regression .....	41
	Multiple Logistic Regression.....	42
<b>12</b>	<b>Trend Tests for Binary Data (106 Patients)</b> .....	43
	Analysis: Multiple Groups Chi-square Test.....	44
	Analysis: Chi-square Test for Trends.....	45
<b>13</b>	<b>Paired Binary (McNemar Test) (139 General Practitioners)</b> .....	47
	Analysis: McNemar’s Test.....	49
<b>14</b>	<b>Multiple Paired Binary Data (Cochran’s Q Test) (139 Patients)</b> .....	51
	Analysis: Cochran’s Q Test.....	52
<b>15</b>	<b>Cox Regression (60 Patients)</b> .....	55
	Simple Cox Regression.....	56
	Multiple Cox Regression .....	58
<b>16</b>	<b>Cox Regression with Time-dependent Variables (60 Patients)</b> .....	59
	Simple Cox Regression.....	61
	Cox Regression with Time-dependent Variables .....	61
<b>17</b>	<b>Validating Qualitative Diagnostic Tests (575 Patients)</b> .....	63
	Validating the Qualitative Diagnostic Test.....	64
<b>18</b>	<b>Validating Quantitative Diagnostic Tests (17 Patients)</b> .....	67
	Validating Quantitative Diagnostic Test .....	67
<b>19</b>	<b>Reliability Assessment of Qualitative Diagnostic Tests (17 Patients)</b> .....	69
	Analysis: Calculate Cohen’s Kappa.....	70

<b>20 Reliability Assessment of Quantitative Diagnostic Tests (17 Patients)</b> .....	71
Intraclass Correlation for Reliability Assessment of Diagnostic Battery .....	71
<b>21 Final Remarks</b> .....	73
<b>Index</b> .....	75

# Chapter 1

## Introduction

This small book contains all statistical tests that are relevant for starters on SPSS. It begins with one sample tests, paired and unpaired basic tests for continuous data, both the parametric and non-parametric ones. Subsequently, paired and unpaired analysis of variance (ANOVA) are addressed, as well as the non-parametric alternatives for these purposes, the Friedman's and Kruskal–Wallis' tests. Then, regression methods are addressed: linear, logistic, and Cox regression. Finally, attention is given to paired/unpaired binary data and multiple paired binary data, trend tests, and diagnostic tests.

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. 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 a CD containing the datafiles of the examples given is made available by the editor.

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 separate data or enter their entire datafile, 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.

The cookbook 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.

The 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 (4th edition, 2009, Springer Dordrecht).

Finally, two last and important points.

1. A datafile 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

## One-Sample Continuous and Binary Data (t-Test, z-Test) (10 and 55 Patients)

### One Sample of Continuous Data

Primary scientific question: is the observed reduction of mean blood pressure after treatment larger than zero reduction.

Variable
3.00
4.00
-1.00
3.00
2.00
-2.00
4.00
3.00
-1.00
2.00

Var = decrease of mean blood pressure after treatment (mmHg) (Var = variable).

### Analysis: One-Sample t-Test

Command:

Analyze – compare means – one sample t-test – test variable: enter Var 00001 – ok.

One-sample test						
Test value = 0						
	t	df	Sig. (2-tailed)	Mean difference	95% confidence interval of the difference	
					Lower	Upper
VAR00001	2.429	9	0.038	1.70000	0.1165	3.2835

The above table shows that the t-value equals 2.429, which means that with  $(10-1) = 9$  degrees of freedom a significant effect is obtained with  $p = 0.038$ . The reduction of mean blood pressure has an average value of 1.7 mmHg, and this average reduction is significantly larger than a reduction of 0 mmHg.

### One Sample of Binary Data

Primary scientific question: is the mean of patients who respond significantly larger than a number of 0.

1 Variable		
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	1	1
0	1	1
0		

Var 1 = responder to antihypertensive drug or not (1 or 0)  
(Var = variable).

### Analysis: z-Test

Command:  
Analyze – descriptive statistics – descriptives – variable(s): enter Var 00001 – options:  
mark: mean, sum, Std. error mean – continue – ok.

Descriptive statistics				
	<u>N</u>	<u>Sum</u>	<u>Mean</u>	
	Statistic	Statistic	Statistic	Std. error
Afdeling	55	20.00	0.3636	0.06546
Valid N (listwise)	55			

The z-value as obtained equals  $0.3636/0.06546 = 5.5545$ . This z-value is much larger than 1.96, and, therefore, the null hypothesis of no difference from a proportion of 0 can be rejected with  $p < 0.001$ . This proportion of 20/55 is significantly different from 0.





# Chapter 3

## Paired Continuous Data (Paired-t, Wilcoxon)

### (10 Patients)

Primary scientific question: is the sleeping pill more efficaceous than the placebo.

Variable					
1	2	3	4	5	6
6.1	5.2	54	0	0	0
7	7.9	55	0	0	0
8.2	3.9	78	1	0	0
7.6	4.7	53	1	1	1
6.5	5.3	49	1	1	1
8.4	5.4	85	1	1	1
6.9	4.2	77	0	1	1
6.7	6.1	66	0	1	1
7.4	3.8	79	0	0	1
5.8	6.3	67	1	0	1

Var 1 = hours of sleep during the use of a sleeping pill (Var = variable).

Var 2 = hours of sleep during placebo.

Var 3 = age.

Var 4 = gender.

Var 5 = co-morbidity.

Var 6 = co-medication.



The sleeping pill performs significantly better than does placebo with a p-value of 0.011, which is much smaller than 0.05. The difference is, thus, highly significant.

## Alternative Analysis: Wilcoxon

If the data do not have a Gaussian distribution, this method will be required, but with Gaussian distributions it may be applied even so.

Command:

Analyze – nonparametric – 2 related samples – further as above (Wilcoxon has already been marked).

Test statistics <sup>a</sup>	
Effect treatment 2 – effect treatment 1	
Z	-2.346 <sup>b</sup>
Asymp. Sig. (2-tailed)	0.019

<sup>a</sup>Wilcoxon Signed Ranks Test

<sup>b</sup>Based on positive ranks

As demonstrated in the above table, also according to the non-parametric Wilcoxon's test the sleeping pill works significantly better than the placebo. The p-value of difference here equals  $p = 0.019$ . This p-value is larger than the p-value of the paired t-test, but still a lot smaller than 0.05, and, so, the effect is still highly significant.



## Chapter 4

# Unpaired Continuous Data (Unpaired t-Tests, Mann–Whitney) (20 Patients)

Primary scientific question: is the sleeping pill more efficacious than the placebo.

Variable				
1	2	3	4	5
0	6	45	0	1
0	7.1	45	0	1
0	8.1	46	0	0
0	7.5	37	0	0
0	6.4	48	0	1
0	7.9	76	1	1
0	6.8	56	1	1
0	6.6	54	1	0
0	7.3	63	1	0
0	5.6	75	0	0
1	5.1	64	1	0
1	8	35	0	1
1	3.8	46	1	0
1	4.4	44	0	1
1	5.2	64	1	0
1	5.4	75	0	1
1	4.3	65	1	1
1	6	84	1	0
1	3.7	35	1	0
1	6.2	46	0	1

Var 1 = group 0 has placebo, group 1 has sleeping pill.

Var 2 = hours of sleep.

Var 3 = age.

Var 4 = gender.

Var 5 = co-morbidity.

### Analysis: Unpaired t-Test

Command:

Analyze – compare means – independent samples t test – dialog box – grouping variable (grouping var 1, test var 2) – defining groups – continue – ok.

		Independent samples test								
Levene's test for equality of variance		t-test for equality of means					95% confidence interval of the difference			
F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	Lower	Upper		
Effect treatment	Equal variances assumed	1.060	0.317	3.558	18	0.002	1.72000	0.48339	0.70443	2.73557
	Equal variances not assumed			3.558	15.030	0.003	1.72000	0.48339	0.68986	2.75014

The unpaired t-test shows that a significant difference exists between the sleeping pill and the placebo with a p-value of 0.002 and 0.003. Generally, it is better to use the largest of the p-values given, because the smallest p-value makes assumptions that are not always warranted, like, for example, in the above table the presence of equal variances of the means of variables 1 and 2.

### Alternative Analysis: Mann–Whitney Test

Just like with the Wilcoxon’s test instead of the t-test for paired data, this test is a non-parametric alternative for the unpaired t-test. If the data have a Gaussian distribution, then it is appropriate to use this test even so. More explanations about Gaussian or parametric distributions are given in the textbook “Statistics Applied to Clinical Trials”, 4th edition, 2009, Springer Dordrecht by the same authors.

Command:

Analyze – nonparametric – two independent samples – further idem unpaired t-test – mark Mann–Whitney (has already been marked) – ok.

Test statistics <sup>a</sup>	
	Effect treatment
Mann–Whitney U	12.500
Wilcoxon W	67.500
Z	–2.836
Asymp. sig. (2-tailed)	0.005
Exact sig. (2 * (1-tailed sig.))	0.300 <sup>b</sup>

<sup>a</sup>Grouping variable: group

<sup>b</sup>Not corrected for ties

The non-parametric Mann–Whitney test produces approximately the same result as the unpaired t-test. The p-value equals 0.005 corrected for multiple identical values and even 0.003 uncorrected. Which of the two is given in the final data report, does not make too much of a difference.





# Chapter 5

## Linear Regression (20 Patients)

Primary scientific question: is the sleeping pill more efficaceous than the placebo.

Variable				
1	2	3	4	5
0.00	6.00	65.00	0.00	1.00
0.00	7.10	75.00	0.00	1.00
0.00	8.10	86.00	0.00	0.00
0.00	7.50	74.00	0.00	0.00
0.00	6.40	64.00	0.00	1.00
0.00	7.90	75.00	1.00	1.00
0.00	6.80	65.00	1.00	1.00
0.00	6.60	64.00	1.00	0.00
0.00	7.30	75.00	1.00	0.00
0.00	5.60	56.00	0.00	0.00
1.00	5.10	55.00	1.00	0.00
1.00	8.00	85.00	0.00	1.00
1.00	3.80	36.00	1.00	0.00
1.00	4.40	47.00	0.00	1.00
1.00	5.20	58.00	1.00	0.00
1.00	5.40	56.00	0.00	1.00
1.00	4.30	46.00	1.00	1.00
1.00	6.00	64.00	1.00	0.00
1.00	3.70	33.00	1.00	0.00
1.00	6.20	65.00	0.00	1.00

- Var 1 = group 0 has placebo, group 1 has sleeping pill.
- Var 2 = hours of sleep.
- Var 3 = age.
- Var 4 = gender.
- Var 5 = co-morbidity.

Similarly to an unpaired t-test, linear regression can be used to test whether there is a significant difference between two treatment modalities. To see how it works picture a linear regression of cholesterol levels and diameters of coronary arteries. It may show that the higher the cholesterol, the narrower the coronary arteries. Cholesterol levels are drawn on the x-axis, coronary diameters on the y-axis, and the best fit

regression line can be about the data can be calculated. Instead of a continuous variable on the x-axis, a binary variable can be adequately used, e.g., two treatment modalities, a bad and a good treatment. With hours of sleep on the y-axis, a nice linear regression analysis can be performed: the better the sleeping treatment, the larger the numbers of sleeping hours. The treatment modality is called the x-variable, or independent variable, or exposure variable, or predictor variable, the hours of sleep is called the y-variable, or dependent variable or outcome variable.

## Simple Linear Regression

Command:

Analyze – regression – linear - dependent = treatment – independent = group – ok – three tables.

Model summary				
Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.643 <sup>a</sup>	0.413	0.380	1.08089

<sup>a</sup>Predictors: (Constant), group

ANOVA <sup>a</sup>						
Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	14.792	1	14.792	12.661	0.002 <sup>b</sup>
	Residual	21.030	18	1.168		
	Total	35.822	19			

<sup>a</sup>Dependent variable: effect treatment

<sup>b</sup>Predictors: (Constant), group

Coefficients <sup>a</sup>						
Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	6.930	0.342		20.274	0.000
	Group	-1.720	0.483	-0.643	-3.558	0.002

<sup>a</sup>Dependent variable: effect treatment

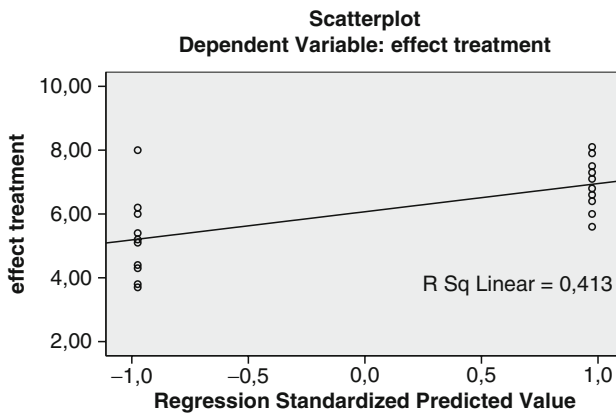
The upper table shows the correlation coefficient ( $R = 0.643 = 64\%$ ).  $R$ -square =  $R^2 = 0.413 = 41\%$ , meaning that, if you know the treatment modality, you will be able to predict the treatment effect (hours of sleep) with 41% certainty. You will then be uncertain with 59% uncertainty.

The magnitude of the  $R$ -value is important for making predictions. However, the size of the study sample is also important: with a sample of say three subjects

little prediction is possible). This is, particularly, assessed in the middle table. It tests with analysis of variance (ANOVA) whether there is a significant correlation between the x and y-variables.

It does so by assessing whether the calculated R-square value is significantly different from an R-square value of 0. The answer is yes. The p-value equals 0.002, and, so, the treatment modality is a significant predictor of the treatment modality.

The bottom table shows the calculated B-value (the regression coefficient). The B-value is obtained by counting/multiplying the individual data values, and it behaves in the regression model as a kind of mean result. Like many mean values from random data samples, this also means, that the B-value can be assumed to follow a Gaussian distribution, and that it can, therefore, be assessed with a t-test. The calculated t-value from these data is smaller than  $-1.96$ , namely  $-3.558$ , and, therefore, the p-value is  $<0.05$ . The interpretation of this finding is, approximately, the same as the interpretation of the R-square value: a significant B-value means that B is significantly smaller (or larger) than 0, and, thus, that the x-variable is a significant predictor of the y-variable. If you square the t-value, and compare it with the F-value of the ANOVA table, then you will observe that the values are identical. The two tests are, indeed, largely similar. One of the two tests is somewhat redundant.



**Fig.** The numbers of hours of sleep during the sleeping pill are larger than during placebo. On the x-axis we have treatment modality, on the y-axis hours of sleep

The above figure shows that the sleeping scores after the placebo are generally smaller than after the sleeping pill. The significant correlation between the treatment modality and the numbers of sleeping hours can be interpreted as a significant difference in treatment efficacy of the two treatment modalities.

Not only treatment modality, but also patient characteristics like age, gender, and co-morbidity may be significant predictors of hours of sleep. The interesting thing about regression analysis is that, in addition to treatment modality, such characteristics can be entered in the model as predictor variables.

## Multiple Linear Regression

Command:

Analyze – regression – linear – dependent = treatment – independent = group and age – ok – three tables.

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.983 <sup>a</sup>	0.966	0.962	0.26684

<sup>a</sup>Predictors: (Constant), age, group

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	34.612	2	17.306	243.045	0.000 <sup>b</sup>
	Residual	1.210	17	0.071		
	Total	35.822	19			

<sup>a</sup>Dependent variable: effect treatment

<sup>b</sup>Predictors: (Constant), age, group

Model		Unstandardized coefficients		Standardized coefficients		Sig.
		B	Std. error	Beta	t	
1	(Constant)	0.989	0.366		2.702	0.015
	Group	-0.411	0.143	-0.154	-2.878	0.010
	Age	0.085	0.005	0.890	16.684	0.000

<sup>a</sup>Dependent variable: effect treatment

In the above multiple regression two predictor variable have been entered: treatment modality and age. The tables resemble strongly the simple linear regression tables. The most important difference is the fact that now the effect of two x-variables is tested simultaneously. The R and the R-square values have gotten much larger, because two predictors, generally, given more information about the y-variable than a single one.  $R\text{-square} = R^2 = 0.966 = 97\%$ , meaning that, if you know the treatment modality and age of a subject from this sample, then you can predict the treatment effect (the numbers of sleeping hours) with 97% certainty, and that you are still uncertain at the amount of 3%.

The middle table takes into account the sample size, and tests whether this R-square value is significantly different from an R-square value of 0.0. The p-value equals 0.001, which means it is true. We can conclude that both variables together significantly predict the treatment effect.

The bottom table now shows, instead of a single one, two calculated B-values (the regression coefficients of the two predictors). They behave like means, and can, therefore, be tested for their significance with two t-tests. Both of them are statistically very significant with p-values of 0.010 and 0.0001. This means that both B-values are significantly larger than 0, and that the corresponding predictors are independent determinants of the y-variable. The older you are, the better you will sleep, and the better the treatment, the better you will sleep.

We can now construct a regression equation for the purpose of making predictions for individual future patients.

$$y = a + b_1x_1 + b_2x_2$$

$$\text{Treatment effect} = 0.99 - 0.41 * \text{group} + 0.085 *$$

Age with the sign \* indicating the sign of multiplication. Thus, a patient of 75 years old with the sleeping pill will sleep for approximately 6.995 h. This is what you can predict with 97% certainty.



# Chapter 6

## Repeated Measures ANOVA, Friedman (10 Patients)

Primary scientific question: do the three different pills produce significantly different magnitudes of the numbers sleeping hours.

Variable				
1	2	3	4	5
6.10	6.80	5.20	55.00	0.00
7.00	7.00	7.90	65.00	0.00
8.20	9.00	3.90	74.00	0.00
7.60	7.80	4.70	56.00	1.00
6.50	6.60	5.30	44.00	1.00
8.40	8.00	5.40	49.00	1.00
6.90	7.30	4.20	53.00	0.00
6.70	7.00	6.10	76.00	0.00
7.40	7.50	3.80	67.00	1.00
5.80	5.80	6.30	66.00	1.00

Var 1 = hours of sleep after pill 1.  
 Var 2 = hours of sleep after pill 2.  
 Var 3 = hours of sleep after pill 3.  
 Var 4 = age.  
 Var 5 = gender.

### Repeated Measurements ANOVA

Command:

Analyze – general linear model – repeated measurements – define factors – within-subjects factor names: treat – number levels: 3 – add – define – enter treat 1, 2, and 3 in box: “Within-subjects Variables” – ok.

Mauchly's test of sphericity<sup>a</sup>

Measure: MEASURE 1

Within subjects effect	Mauchly's W	Aprox. chi-square	df	Sig.	Epsilon <sup>b</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Treat	0.096	18.759	2	0.000	0.525	0.535	0.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

<sup>a</sup>Design: Intercept. Within Subjects Design: treat

<sup>b</sup>May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of within-subjects effects table

Tests of within-subjects effects

Measure: MEASURE 1

Source		Type III sum of squares	df	Mean square	F	Sig.
Treat	Sphericity assumed	24.056	2	12.028	10.639	0.001
	Greenhouse-Geisser	24.056	1.050	22.903	10.639	0.009
	Huynh-Feldt	24.056	1.070	22.489	10.639	0.008
	Lower-bound	24.056	1.000	24.056	10.639	0.010
Error(treat)	Sphericity assumed	20.351	18	1.131		
	Greenhouse-Geisser	20.351	9.453	2.153		
	Huynh-Feldt	20.351	9.627	2.114		
	Lower-bound	20.351	9.000	2.261		

The repeated measures ANOVA tests whether a significant difference exists between three treatments. An important criterion for validity of the test is the presence of sphericity in the data, meaning that all data come from Gaussian distributions. It appears from the above upper table that this is not true, because based on this table we are unable to reject the null-hypothesis of non-sphericity. This means that an ANOVA test corrected for non-sphericity has to be performed. There are three possibilities: the Greenhouse, Huynh, and Lower-bound methods. All of them produce a much larger p-value than the uncorrected method, but the result is still statistically highly significant with p-values of 0.009, 0.008, and 0.010. A significant difference between the treatments has, thus, been demonstrated. However, we do not yet know whether the significant difference is located between the treatments 1 and 2, between the treatments 1 and 3, or between the treatments 2 and 3. In order to find out three separate paired t-tests have to be performed. Note, that with multiple t-tests it is better to reduce the cut-off level for statistical significance to approximately 0.01 (more information about the adjustments for multiple testing including the Bonferroni procedure is given in the textbook "Statistics Applied to Clinical Trials", 4<sup>th</sup> edition, 2009, Springer Dordrecht by the same author).



## Alternative: Friedman Test (Better with Non-Gaussian Distributions)

Command:

Analyze – nonparametric – k-related samples – test variables: treat 1, treat 2, treat 3 – mark: Friedman – ok.

Test statistics <sup>a</sup>	
N	10
Chi-square	7.579
df	2
Asymp. sig.	0.023

<sup>a</sup>Friedman test

The result is significant but the p-value is markedly larger than the p-value of the ANOVA, i.e., 0.023. Just like with the above ANOVA we will have to perform additional tests to determine where the difference of the three treatments is located.

For that purpose three Wilcoxon's tests could be performed (and adjustment for multiple testing can be done similarly to the above procedure: using either a p-value of 0.01 or a Bonferroni adjustment, see textbook "Statistics Applied to Clinical Trials", 4th edition, 2009, Springer Dordrecht by the same author).



## Chapter 7

### Mixed Models (20 Patients)

Primary scientific question: is there a significant difference in efficacy of treatment 0 and treatment 1 after adjustment for five repeated measurements.

Variable							
1	2	3	4	1	2	3	4
Patient	Week	Outcome	Treatment	Patient	Week	Outcome	Treatment
1.00	1.00	1.66	0.00	6.00	4.00	1.87	0.00
1.00	2.00	1.62	0.00	6.00	5.00	1.88	0.00
1.00	3.00	1.57	0.00	7.00	1.00	2.04	0.00
1.00	4.00	1.52	0.00	7.00	2.00	2.06	0.00
1.00	5.00	1.50	0.00	7.00	3.00	1.95	0.00
2.00	1.00	1.69	0.00	7.00	4.00	1.90	0.00
2.00	2.00	1.71	0.00	7.00	5.00	1.91	0.00
2.00	3.00	1.60	0.00	8.00	1.00	2.07	0.00
2.00	4.00	1.55	0.00	8.00	2.00	2.09	0.00
2.00	5.00	1.56	0.00	8.00	3.00	1.98	0.00
3.00	1.00	1.92	0.00	8.00	4.00	1.93	0.00
3.00	2.00	1.94	0.00	8.00	5.00	1.94	0.00
3.00	3.00	1.83	0.00	9.00	1.00	2.30	0.00
3.00	4.00	1.78	0.00	9.00	2.00	2.32	0.00
3.00	5.00	1.79	0.00	9.00	3.00	2.21	0.00
4.00	1.00	1.95	0.00	9.00	4.00	2.16	0.00
4.00	2.00	1.97	0.00	9.00	5.00	2.17	0.00
4.00	3.00	1.86	0.00	10.00	1.00	2.36	0.00
4.00	4.00	1.81	0.00	10.00	2.00	2.35	0.00
4.00	5.00	1.82	0.00	10.00	3.00	2.26	0.00
5.00	1.00	1.98	0.00	10.00	4.00	2.23	0.00
5.00	2.00	2.00	0.00	10.00	5.00	2.20	0.00
5.00	3.00	1.89	0.00	11.00	1.00	1.57	1.00
5.00	4.00	1.84	0.00	11.00	2.00	1.82	1.00
5.00	5.00	1.85	0.00	11.00	3.00	1.83	1.00
6.00	1.00	2.01	0.00	11.00	4.00	1.83	1.00
6.00	2.00	2.03	0.00	11.00	5.00	1.82	1.00
6.00	3.00	1.92	0.00	12.00	1.00	1.60	1.00

(continued)

Variable (continued)							
1	2	3	4	1	2	3	4
Patient	Week	Outcome	Treatment	Patient	Week	Outcome	Treatment
12.00	2.00	1.85	1.00	16.00	4.00	2.21	1.00
12.00	3.00	1.89	1.00	16.00	5.00	2.17	1.00
12.00	4.00	1.89	1.00	17.00	1.00	1.95	1.00
12.00	5.00	1.85	1.00	17.00	2.00	2.20	1.00
13.00	1.00	1.83	1.00	17.00	3.00	2.25	1.00
13.00	2.00	2.08	1.00	17.00	4.00	2.24	1.00
13.00	3.00	2.12	1.00	17.00	5.00	2.20	1.00
13.00	4.00	2.12	1.00	18.00	1.00	1.98	1.00
13.00	5.00	2.08	1.00	18.00	2.00	2.23	1.00
14.00	1.00	1.86	1.00	18.00	3.00	2.28	1.00
14.00	2.00	2.11	1.00	18.00	4.00	2.27	1.00
14.00	3.00	2.16	1.00	18.00	5.00	2.24	1.00
14.00	4.00	2.15	1.00	19.00	1.00	2.21	1.00
14.00	5.00	2.11	1.00	19.00	2.00	2.46	1.00
15.00	1.00	2.80	1.00	19.00	3.00	2.57	1.00
15.00	2.00	2.14	1.00	19.00	4.00	2.51	1.00
15.00	3.00	2.19	1.00	19.00	5.00	2.48	1.00
15.00	4.00	2.18	1.00	20.00	1.00	2.34	1.00
15.00	5.00	2.14	1.00	20.00	2.00	2.51	1.00
16.00	1.00	1.92	1.00	20.00	3.00	2.55	1.00
16.00	2.00	2.17	1.00	20.00	4.00	2.55	1.00
16.00	3.00	2.22	1.00	20.00	5.00	2.52	1.00

Var 00001 = patient number (Var = variable).

Var 00002 = week of treatment (1–5).

Var 00003 = outcome (HDL cholesterol).

Var 00004 = treatment modality (0 or 1).

It is appropriate, whenever possible, to use a summary estimate of repeated data. For example, the area under the curve of drug concentration–time curves is used in clinical pharmacology as an estimate of bioavailability of a drug. Also, maximal values, mean values, changes from baseline are applied for the same purpose. The disadvantage of this approach is that it does not use the data fully, because summary measures are used instead of the individual data, and, therefore, precision may be lost, but, otherwise, the approach is unbiased, and can be used perfectly well. As an alternative and more precise method a mixed-linear model, as reviewed below, can be used.

In the above table an example is given of a parallel-group study of the effect of two statins on HDL cholesterol. HDL cholesterol is measured every week for 5 weeks. The averages of the five repeated measures in one patient are calculated and an unpaired t-test was used to compare these averages in the two treatment groups. The overall average in group 0 was 1.925 (SEM 0.0025), in group 1 2.227 (SE 0.227). With 18 degrees of freedom and a t-value of 1.99 the difference did not obtain statistical significance,  $0.05 < p < 0.10$ . There seems to be, expectedly, a strong positive correlation between the five repeated measurements in one patient. In order to take account of this strong positive correlation a random-effects mixed-linear model is used.

For this particular analysis all measurements have to be ordered in a single column, not in five columns side by side. In a second column the time of the separate measurements have to be noted.

## Mixed Effects Analysis

Command:

Analyze – mixed models – linear – specify subjects and repeated – continue – linear mixed model – dependent: var 00003 – factors: var 00001, var 00002, var 00004 – fixed – build nested term – var 00004 – add – var 000002 – add – var 00002 build term by\* var 00004 – var 00004 \* var 00002 – add – continue – ok.

Type III tests of fixed effects<sup>a</sup>

Source	Numerator df	Denomination df	F	Sig.
Intercept	1	90	6988.626	.000
VAR00004	1	90	20.030	.000
VAR00002	4	90	.377	.825
VAR00002 * VAR00004	4	90	1.603	.181

<sup>a</sup>Dependent variable: VAR00003

The above table shows that the treatment modality (var 00004) after adjustment for the repeated nature of the data is a significant predictor of HDL cholesterol levels (var 00003) with  $p < 0.0001$ . Treatment 1 performs significantly better than does treatment 0.



## Chapter 8

### One-Way-ANOVA, Kruskal–Wallis (30 Patients)

Primary scientific question: do the two sleeping pills and the placebo produce significantly different magnitudes of numbers of sleeping hours.

Variable				
1	2	3	4	5
0.00	6.00	45.00	0.00	1.00
0.00	7.10	45.00	0.00	1.00
0.00	8.10	46.00	0.00	0.00
0.00	7.50	37.00	0.00	0.00
0.00	6.40	48.00	0.00	1.00
0.00	7.90	76.00	1.00	1.00
0.00	6.80	56.00	1.00	1.00
0.00	6.60	54.00	1.00	0.00
0.00	7.30	63.00	1.00	0.00
0.00	5.60	75.00	0.00	0.00
1.00	5.10	64.00	1.00	0.00
1.00	8.00	35.00	0.00	1.00
1.00	3.80	46.00	1.00	0.00
1.00	4.40	44.00	0.00	1.00
1.00	5.20	64.00	1.00	0.00
1.00	5.40	75.00	0.00	1.00
1.00	4.30	65.00	1.00	1.00
1.00	6.00	84.00	1.00	0.00
1.00	3.70	35.00	1.00	0.00
1.00	6.20	46.00	0.00	1.00
2.00	4.10	43.00	0.00	0.00
2.00	7.00	56.00	0.00	0.00
2.00	2.80	65.00	0.00	0.00
2.00	3.40	66.00	0.00	1.00
2.00	4.20	74.00	1.00	1.00
2.00	4.40	56.00	1.00	1.00
2.00	3.30	45.00	0.00	1.00
2.00	5.00	47.00	1.00	1.00
2.00	2.70	65.00	0.00	1.00
2.00	5.20	56.00	1.00	0.00

Var 1 = group 0 has placebo, group 1 has sleeping pill 1, group 2 sleeping pill 2 (Var = variable).

Var 2 = hours of sleep.

Var 3 = age.

Var 4 = gender.

Var 5 = co-morbidity.

## One-Way-ANOVA

Command:

Analyze – compare means – one-way anova – dependent lists: effect treat – factor – group – ok.

ANOVA					
Effect treatment	Sum of squares	df	Mean square	F	Sig.
Between groups	37.856	2	18.928	14.110	0.000
Within groups	36.219	27	1.341		
Total	74.075	29			

A significant difference between the three treatments has been demonstrated with a p-value of 0.0001. Like with the paired data of the previous chapter the conclusion is drawn: a difference exists, but we don't yet know whether the difference is between treatments 1 and 2, 2 and 3, or 1 and 3. Three subsequent unpaired t-tests are required to find out. Similarly to the tests of Chapter 5, a smaller p-value for rejecting the null-hypothesis is recommended, for example, 0.01 instead of 0.05. This is, because with multiple testing the chance of type 1 errors of finding a difference where there is none is enlarged, and this chance has to be adjusted.

Like the Friedman test can be applied for comparing three or more paired samples as a non-Gaussian alternative to the paired ANOVA test (see Chapter 6), the Kruskal–Wallis test can be used as a non-Gaussian alternative to the above unpaired ANOVA test. It is used for the analysis of the above data.

## Alternative Test: Kruskal–Wallis Test (Better with Non-Gaussian Distributions)

Command:

Analyze – nonparametric – legacy dialogs – k independent samples – test variable list: effect treat – group variable: var 00001(??) – Define range – minimum: 0 – maximum: 2 – continue – mark: Kruskal–Wallis – ok.



Test statistics <sup>a,b</sup>	
	Effect treatment
Chi-square	15.171
df	2
Asymp. Sig.	0.001

<sup>a</sup>Kruskal–Wallis test

<sup>b</sup>Grouping variable: group

The Kruskal–Wallis test is significant with a p-value of no less than 0.001. This means that the three treatments are very significantly different from one another. We don't know, just like with the above unpaired ANOVA, where the difference is. The advice is to perform three additional Mann–Whitney tests to find out whether the difference is between the treatments 1 and 2, 2 and 3, or 1 and 3. Again, a subsequent reduction of the p-value or a Bonferroni test is appropriate.



# Chapter 9

## Trend Test for Continuous Data (30 Patients)

Primary scientific question: do incremental treatment dosages cause incremental beneficial effect on blood pressure.

Variable	
1	2
1.00	122.00
1.00	113.00
1.00	131.00
1.00	112.00
1.00	132.00
1.00	114.00
1.00	130.00
1.00	115.00
1.00	129.00
1.00	122.00
2.00	118.00
2.00	109.00
2.00	127.00
2.00	110.00
2.00	126.00
2.00	111.00
2.00	125.00
2.00	112.00
2.00	124.00
2.00	118.00
3.00	115.00
3.00	105.00
3.00	125.00
3.00	106.00
3.00	124.00
3.00	107.00
3.00	123.00
3.00	108.00
3.00	122.00
3.00	115.00

Var 1 = treatment dosage (Var = variable).  
 Var 2 = treatment response (mean blood pressure after treatment).

## Trend Analysis for Continuous Data

We first perform a one-way ANOVA (see also Chapter 8) to see if there are any significant differences in the data. If not, we will perform a trend test using simple linear regression.

Command:

Analyze – compare means – One-way ANOVA – dependent list: mean blood pressure after treatment – factor: treatment dosage – ok.

VAR00002		ANOVA			
Model	Sum of squares	df	Mean square	F	Sig.
Between groups	246.667	2	123.333	2.035	0.150
Within groups	1636.000	27	60.593		
Total	1882.667	29			

The above table shows that there is no significant difference in efficacy between the treatment dosages, and so, sadly, this is a negative study. However, a trend test having just 1 degree of freedom has more sensitivity than a usual one-way ANOVA, and it could, therefore, be statistically significant even so.

Command:

Analyze – regression – linear – dependent = mean blood pressure after treatment – independent = treatment dosage – ok.

ANOVA <sup>a</sup>						
Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	245.000	1	245.000	4.189	0.050 <sup>b</sup>
	Residual	1637.667	28	58.488		
	Total	1882.667	29			

<sup>a</sup>Dependent variable: VAR00002

<sup>b</sup>Predictors: (Constant), VAR00001

Four tables are given, we will only use the third one as shown above. The table shows that treatment dosage is a significant predictor of treatment response with a p-value of 0.050. There is, thus, a significantly incremental response with incremental dosages.

# Chapter 10

## Unpaired Binary Data (Chi-Square, Crosstabs)

### (55 Patients)

Primary scientific question: is there a significant difference between the risks of falling out of bed at the departments of surgery and internal medicine.

		Fall out of bed		
		No	Yes	
Number of patient	Department surgery	0	20	15
	Internal department	1	5	15

The above consistency table of the data shows that at both departments the same numbers of patients fall out of bed. However, at the department of surgery many more patients do not fall out of bed than at the internal department.

Variable				
1	2	3	4	5
0	1	50	0	1
0	1	76	0	1
0	1	57	1	1
0	1	65	0	1
0	1	46	1	1
0	1	36	1	1
0	1	98	0	0
0	1	56	1	0
0	1	44	0	0
0	1	76	1	1
0	1	75	1	1
0	1	74	1	1
0	1	87	0	0
0	1	45	0	0
0	1	46	1	0
0	0	47	0	0
0	0	48	1	0
0	0	87	0	0
0	0	65	0	0

(continued)

Variable (continued)				
1	2	3	4	5
0	0	50	1	0
0	0	49	0	0
0	0	58	1	1
0	0	93	0	0
0	0	84	1	1
0	0	57	1	1
0	0	48	0	0
0	0	35	0	0
0	0	26	1	0
0	0	76	1	0
0	0	56	0	0
0	0	56	0	0
0	0	35	1	0
0	0	29	0	0
0	0	78	0	0
0	0	67	0	0
1	1	54	1	0
1	1	65	1	0
1	1	74	1	1
1	1	73	0	1
1	1	85	0	1
1	1	65	0	1
1	1	74	1	1
1	1	65	1	1
1	1	75	1	1
1	1	45	0	1
1	1	67	1	1
1	1	76	0	0
1	1	65	1	0
1	1	86	0	0
1	1	76	1	0
1	0	95	0	0
1	0	46	1	0
1	0	57	0	0
1	0	46	1	0
1	0	78	0	1

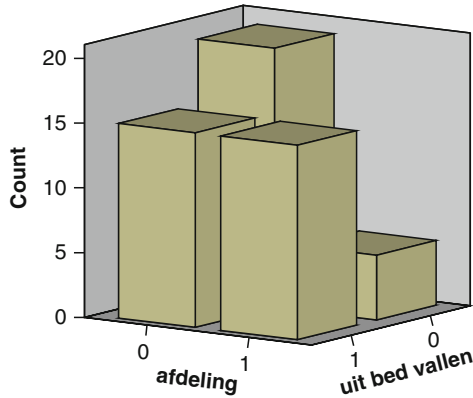
Var 1 = 0 = dept surgery, 1 = internal dept  
(Var = variable).

Var 2 = falling out of bed.

Var 3 = age.

Var 4 = gender.

Var 5 = letter of complaint, yes or no.



**Fig.** Command: graphs – legacy dialogs – 3D charts – x-axis groups – z-axis groups –define – x-axis dept – z-axis falling out of bed – ok

At both departments approximately the same number of patients falls out of bed. However, at department-0 many more patients do not fall out of bed than at department-1.

### Analysis: Chi-square Test

Command:

Analyze – descriptive statistics – crosstabs – rows var 1 – columns var 2 – statistics – chi-square – ok.

Chi-square tests					
	Value	df	Asymp. sig. (2-sided)	Exact sig. (2-sided)	Exact sig. (1-sided)
Pearson chi-square	5.304 <sup>b</sup>	1	0.021		
Continuity correction <sup>a</sup>	4.086	1	0.043		
Likelihood ratio	5.494	1	0.019		
Fisher's exact test				0.027	0.021
Linear-by-linear association	5.207	1	0.022		
No. of valid cases	55				

<sup>a</sup>Computed only for a 2 × 2 table

<sup>b</sup>0 cells (.0%) have expected count less than 5. The minimum expected count is 9.09

The chi-square test (Pearson chi-square) shows that a significant difference between the surgical and internal departments exists in patterns of patients falling out of bed. The p-value equals 0.021, and this is much smaller than 0.05. Several contrast tests are given in the above results table. They produce approximately similar p-values. This supports the accuracy of the chi-square test for these data.





## Chapter 11

### Logistic Regression (55 Patients)

Primary scientific question: Primary scientific question: is there a significant difference between the risks of falling out of bed at the departments of surgery and internal medicine.

		Fall out of bed		
		No	Yes	
Number of patients	Surgical department	0	20	15
	Internal department	1	5	15

The above consistency table of the data shows that at both departments the same numbers of patients fall out of bed. However, at the department of surgery many more patients do not fall out of bed than at the internal department.

Variable				
1	2	3	4	5
0	1	50	0	1
0	1	76	0	1
0	1	57	1	1
0	1	65	0	1
0	1	46	1	1
0	1	36	1	1
0	1	98	0	0
0	1	56	1	0
0	1	44	0	0
0	1	76	1	1
0	1	75	1	1
0	1	74	1	1
0	1	87	0	0
0	1	45	0	0
0	1	46	1	0
0	0	47	0	0
0	0	48	1	0

(continued)

Variable (continued)				
1	2	3	4	5
0	0	87	0	0
0	0	65	0	0
0	0	50	1	0
0	0	49	0	0
0	0	58	1	1
0	0	93	0	0
0	0	84	1	1
0	0	57	1	1
0	0	48	0	0
0	0	35	0	0
0	0	26	1	0
0	0	76	1	0
0	0	56	0	0
0	0	56	0	0
0	0	35	1	0
0	0	29	0	0
0	0	78	0	0
0	0	67	0	0
1	1	54	1	0
1	1	65	1	0
1	1	74	1	1
1	1	73	0	1
1	1	85	0	1
1	1	65	0	1
1	1	74	1	1
1	1	65	1	1
1	1	75	1	1
1	1	45	0	1
1	1	67	1	1
1	1	76	0	0
1	1	65	1	0
1	1	86	0	0
1	1	76	1	0
1	0	95	0	0
1	0	46	1	0
1	0	57	0	0
1	0	46	1	0
1	0	78	0	1

Var 1 = 0 = surgical department,  
 1 = internal department (Var = variable).  
 Var 2 = falling out of bed yes or no.  
 Var 3 = age.  
 Var 4 = gender.  
 Var 5 = letter of complaint yes or no

Similarly to chi-square test, logistic regression can be used to test whether there is a significant difference between two treatment modalities. To see how it works review the linear regression example from Chapter 5. The linear regression model with treatment modality as independent variable (x-variable), and hours of sleep as dependent variable (y-variable = outcome variable) showed that the treatment modality was a significant predictor of the hours of sleep, and, thus, that there was a significant difference between the two treatments. In the current example we have a largely similar situation. The type department is assumed to predict the risk of falling out of bed, and is defined as a binary x-variable. The risk of falling out of bed is the y-variable, but, unlike hours of sleep like in Chapter 5, falling out of bed is not a continuous variable, but rather a binary variable: you either fall or you don't. With binary y-variables linear regression is impossible, and logistic regression is required. Otherwise, the analysis and interpretation is similar to that of the linear regression.

## Simple Logistic Regression

Command:

Analyze – regression – binary logistic – dependent = fall out of bed – covariate = hospital dept – ok – look at “variables in equation”.

		Variables in the equation					
		B	Std. Error	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	VAR00001	1.386	0.619	5.013	1	0.025	4.000
	Constant	-0.288	0.342	0.709	1	0.400	0.750

<sup>a</sup>Variable(s) entered on step 1: VAR00001

The above results table of the logistic regression shows that B (the regression coefficient) for the variable 00001 (which is the hospital department) is a significant predictor of the chance of falling out of bed with a p-value of 0.025. This is a p-value largely similar to that of the chi-square test from Chapter 9. The meaning of this logistic regression is also largely the same as that of the chi-square test. A nice thing with logistic regression is that, unlike with chi-square tests, an odds ratio is given.

The odds ratio equals approximately 4 which can be interpreted as follows: the chance of falling out of bed is about four times larger at the department of surgery than it is at the department of internal medicine.

The significant correlation between the type of department and the risk of falling out of bed can be interpreted as a significant difference in safety at the two departments.

Not only type of department, but also patient characteristics like age, gender, and co-morbidity may be significant predictors of falling out of bed. The interesting thing about regression analysis is that, in addition to treatment modality, such characteristics can be entered in the model as predictor variables.

## Multiple Logistic Regression

First we will test whether age is a significant predictor of falling out of bed.

Command:

Analyze – regression – binary logistic – dependent = fall out of bed – covariate = age – ok – look at “variables in equation”.

		Variables in the equation					
		B	Std. Error	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	VAR00003	0.106	0.027	15.363	1	0.000	1.112
	Constant	-6.442	1.718	14.068	1	0.000	0.002

<sup>a</sup> Variable(s) entered on step 1: VAR00003

The correct conclusion is, indeed, that this is true. Var 00003, age, is an independent determinant with a p-value of 0.0001. The odds ratio equals 1.112, which indicates that each year the chance of falling out of bed increases by 1.112. Subsequently, we will test whether something special is going on: one of the predictors might be a confounder of the other predictor, and, also, interaction between age and department could very well exist. Therefore, we perform a multiple logistic regression with both predictors as x-variables and with an interaction-variable of the two as a third predictor.

Command:

Analyze – regression – binary logistic – dependent = fall out of bed – covariate = hospital dept , age, interaction variable (keep pressing CTRL – age and hospital dept turn blue – then press <a \* b> – ok – look at “variables in equation”.

		Variables in the equation					
		B	Std. error	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	VAR00001 by VAR00003	4.619	736.214	0.000	1	0.995	101.354
	VAR00001	-279.342	44,876.713	0.000	1	0.995	0.000
	VAR00003	0.072	0.025	8.176	1	0.004	1.075
	Constant	-4.577	1.584	8.355	1	0.004	0.010

<sup>a</sup> Variable(s) entered on step 1: VAR00001\* VAR00003, VAR00001, VAR00003

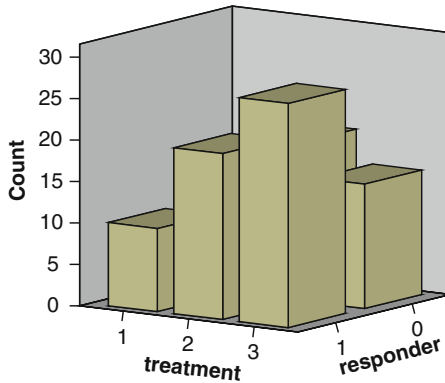
The analysis shows that interaction is not observed, and that the significant effect of the department has disappeared, while age as single variable is a highly significant predictor of falling out of bed with a p-value of 0.004 and an odds ratio of 1.075 per year. The initial significant effect of the difference in department is not caused by a real difference, but rather by the fact that at one department many more elderly patients are admitted than at the other department. After adjustment for age the significant effect of the department has disappeared.



The underneath contingency table shows that with incremental dosages the odds of responding is growing from 0.67 to 1.80.

	Dosage 1	Dosage 2	Dosage 3
Numbers responders	10	20	27
Numbers non-responders	15	19	15
Odds of responding	0.67 (10/15)	1.11 (20/19)	1.80 (27/15)

First, we try and summarize the data in a graph.



**Fig.** Command: graphs – 3D charts – x-axis treatment – z-axis responder – define – x-axis treatment – z-axis responder – ok

The incremental treatment dosages of an antihypertensive drug seem to cause incremental numbers of responders (patients becoming normotensive). However, the numbers of non-responders are the controls and their pattern is equally important.

We, first will perform a multiple groups chi-square test in order to find out whether there is any significant difference in the data.

### Analysis: Multiple Groups Chi-square Test

Command:

Analyze – descriptive statistics – crosstabs – rows var 1 – columns var 2 – statistics – chi-square – ok.

	Chi-square tests		
	Value	df	Asymp. sig. (2-sided)
Pearson chi-square	3.872 <sup>a</sup>	2	0.144
Likelihood ratio	3.905	2	0.142
Linear-by-linear association	3.829	1	0.050
No. of valid cases	106		

<sup>a</sup>0 cells (.0%) have expected count less than 5.  
The Minimum expected count is 11.56

The above table shows that, indeed, the Pearson chi-square value for multiple groups testing is not significant with a chi-square value of 3.872 and a p-value of 0.144, and we have to conclude that there is, thus, no significant difference between the odds of responding to the three dosages.

### Analysis: Chi-square Test for Trends

Subsequently, a chi-square test for trends can be executed, a test that essentially assesses whether the above odds of responding (number of responder/numbers of non-responders per treatment group) increase significantly. The “linear-by-linear association” from the same table is appropriate for the purpose. It has approximately the same chi-square value, but it has only 1 degree of freedom, and, therefore it reaches statistical significance with a p-value of 0.050. There is, thus, a significant incremental trend of responding with incremental dosages.

	Chi-square tests		
	Value	df	Asymp. sig. (2-sided)
Pearson chi-square	3.872 <sup>a</sup>	2	0.144
Likelihood ratio	3.905	2	0.142
Linear-by-linear association	3.829	1	0.050
No. of valid cases	106		

<sup>a</sup>0 cells (.0%) have expected count less than 5.  
The Minimum expected count is 11.56

The trend in this example can also be tested using logistic regression with responding as outcome variable and treatment as independent variable (enter the latter as covariate, not categorical variable).





## Chapter 13

### Paired Binary (McNemar Test)

#### (139 General Practitioners)

Primary scientific question: is there a significant difference between the numbers of practitioners who give life style treatment in the periods before and after education.

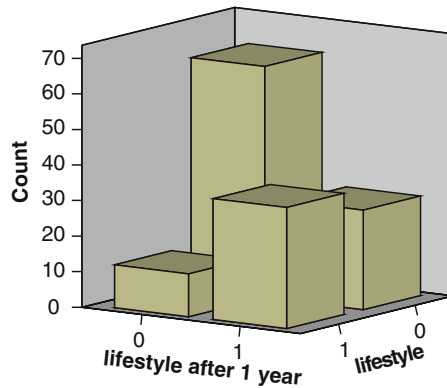
Numbers of practitioners giving life style treatment after education				
		No		
		Yes		
		0		
		1		
Life style treatment	No	0	65	28
before education	Yes	1	12	34

The above table summarizes the numbers of practitioners giving life style treatment in the periods before and after education. Obviously, before education  $65 + 28 = 93$  did not give life style, while after education this number fell to 77. It looks as though the education was somewhat successful.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15					
0	0	59	0	0	0	52	1	1	0	1	46	0	0	1	1	35	0	1	
0	0	56	0	1	0	0	76	1	0	0	1	48	1	1	1	1	45	0	1
0	0	78	0	1	0	0	66	0	0	0	1	51	0	1	1	1	27	1	0
0	0	66	0	1	0	0	67	1	0	0	1	76	1	1	1	1	65	1	1
0	0	55	0	1	0	0	65	0	1	0	1	71	0	1	1	1	54	1	0
0	0	53	0	0	0	0	68	1	1	0	1	76	0	0	1	1	83	1	1
0	0	67	1	0	0	0	76	0	1	0	1	65	0	0	1	1	54	0	0
0	0	58	1	1	0	0	89	1	0	0	1	74	1	0	1	1	76	0	1
0	0	67	1	1	0	0	49	1	0	0	1	65	1	0	1	1	48	0	0
0	0	45	0	1	0	0	55	0	0	0	1	56	1	1	1	1	65	0	1
0	0	87	1	1	0	0	65	0	1	0	1	47	0	0	1	1	46	1	1
0	0	96	0	1	0	0	63	0	1	0	1	38	0	1	1	1	37	0	1
0	0	67	1	0	0	0	62	1	0	0	1	29	0	0	1	1	28	1	0
0	0	54	0	0	0	0	61	1	1	0	1	54	1	1	1	1	76	0	1
0	0	67	1	0	0	0	37	1	0	0	1	63	0	0	1	1	46	1	0
0	0	55	0	1	0	0	39	0	1	0	1	73	1	1	1	1	53	1	1
0	0	86	0	1	0	0	19	0	0	0	1	76	0	0	1	1	86	0	0
0	0	84	0	1	0	0	53	1	1	0	1	85	0	1	1	1	75	0	1
0	0	63	0	1	0	0	78	1	0	0	1	83	1	0	1	1	64	1	0
0	0	61	1	0	0	0	54	0	0	0	1	64	0	0	1	1	53	1	1
0	0	43	0	0	0	0	36	1	1	0	1	53	1	0	1	1	46	0	0
0	0	52	1	0	0	0	54	0	1	0	1	64	0	1	1	1	37	0	0
0	0	62	0	0	0	0	64	1	0	0	1	57	1	1	1	1	49	1	0
0	0	56	0	1	0	0	63	0	0	1	0	56	0	1	1	1	88	1	0
0	0	47	0	1	0	0	58	0	0	1	0	45	0	1	1	1	77	0	1
0	0	38	0	1	0	0	46	0	1	1	0	75	1	0	1	1	68	0	0
0	0	54	1	1	0	0	54	1	1	1	0	84	1	0	1	1	57	1	1
0	0	55	1	1	0	0	36	1	1	1	0	73	1	1	1	1	86	1	0
0	0	77	1	0	0	0	64	0	0	1	0	64	0	1	1	1	75	0	1
0	0	86	1	0	0	0	45	0	0	1	0	53	1	0	1	1	64	0	0
0	0	77	0	1	0	1	52	1	1	1	0	64	0	0	1	1	54	1	1
0	0	45	0	1	0	1	66	1	1	1	0	85	1	1	1	1	55	1	1
0	0	45	0	0	0	1	79	1	1	1	0	67	1	1	1	1	67	0	0
0	0	55	1	0	0	1	53	1	0	1	0	47	0	0	1	1	53	0	0
0	0	53	0	1	0	1	55	0	0	1	0	49	0	0	1	1	49	0	0

- Var 1 = life style treatment before post-graduate education (yes or no) (Var = variable).
- Var 2 = life style treatment after post-graduate education (yes or no).
- Var 3 = age.
- Var 4 = gender.
- Var 5 = co-morbidity.

We will start with a graph of the data.



**Fig.** Command: graphs – legacy dialogs – 3D charts – x-axis groups – z-axis groups – define – x-axis life style after education – z-axis life style before education – ok

The paired observations show that twice no life style treatment was given by 65 practitioners, twice yes life style treatments by 34 practitioners. Furthermore, 28 practitioners started to give life style treatment after postgraduate education, while, in contrast, 12 stopped giving life style treatment after the education program. McNemar's test is used to statistically test the differences.

### Analysis: McNemar's Test

Command:

Analyze – nonparametric – 2 samples related – current selection – var 1 = var 1 – var 2 = var 2 – test-pair list – McNemar – ok.

Test statistics <sup>a</sup>	
Lifestyle and lifestyle after 1 year	
N	139
Chi-square <sup>b</sup>	5.625
Asymp. sig.	0.018

<sup>a</sup>McNemar test

<sup>b</sup>Continuity corrected

The above test is statistically significant with a p-value of 0.018, which is a lot smaller than 0.05. The conclusion is drawn that a real difference between the numbers of practitioners giving life style treatment after and before education is observed. The postgrade education has, obviously, been helpful.



# Chapter 14

## Multiple Paired Binary Data (Cochran's Q Test)

### (139 Patients)

Scientific question: is there a significant difference between the numbers of responders who have been treated differently three times.

Variables											
1	2	3	1	2	3	1	2	3	1	2	3
0	0	0	0	1	0	0	1	0	1	0	0
0	0	1	0	0	1	0	0	0	1	0	0
0	0	0	0	1	1	0	0	0	1	0	1
0	0	1	0	1	0	0	1	1	1	0	1
0	0	1	0	0	0	0	0	1	1	1	0
0	0	1	0	1	0	0	1	1	1	1	1
0	0	1	0	0	1	0	0	1	1	1	0
0	0	0	0	1	1	0	0	0	1	1	1
0	1	0	0	0	1	0	0	0	1	0	0
0	1	1	0	1	0	0	1	0	1	0	1
0	1	1	0	1	0	0	1	0	1	0	0
0	0	1	0	0	0	0	1	1	1	0	1
0	1	1	0	0	1	0	0	0	1	1	1
0	0	1	0	0	1	0	0	1	1	1	1
0	0	1	0	0	1	0	0	1	1	0	1
0	0	0	0	1	1	0	1	1	1	0	1
0	1	0	0	1	0	0	0	0	1	1	0
0	0	1	0	0	1	0	1	1	1	1	1
0	0	1	0	0	0	0	0	0	1	0	0
0	0	1	0	1	1	0	0	1	1	0	1
0	0	1	0	1	1	0	0	1	1	0	1
0	0	1	0	1	0	0	1	0	1	1	0
0	1	0	0	0	0	0	0	0	1	1	1
0	1	0	0	1	1	0	1	0	1	0	0
0	1	0	0	0	1	0	0	1	1	0	0
0	0	0	0	1	0	0	1	1	1	1	0
0	0	1	0	0	0	1	0	1	1	1	0
0	0	1	0	0	0	1	0	1	1	0	1
0	0	1	0	0	1	1	1	0	1	0	0
0	1	1	0	1	1	1	1	0	1	1	1
0	1	1	0	1	1	1	1	1	1	1	0
0	1	0	0	0	0	1	0	1	1	0	1
0	1	0	0	0	0	1	1	0	1	0	0
0	0	1	0	1	1	1	0	0	1	1	1
0	0	1	0	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	1	1	0	0
1	0	0	1	0	0	1	1	0	0	0	0

Var 1 = responder to treatment 1 (yes or no, 1 or 0) (Var = variable).  
 Var 2 = responder to treatment 2.  
 Var 3 = responder to treatment 3.

The above table shows three paired observations in one patient. The paired property of these observations has to be taken into account because of the, generally, positive correlation between paired observations. Cochran’s Q test is appropriate for that purpose.

### Analysis: Cochran’s Q Test

Command:

Analyze – nonparametric tests – k related samples – mark: Cochran’s Q – test variables: treatment 1, treatment 2, treatment 3 – ok.

Test statistics	
N	139
Cochran’s Q	10.133 <sup>a</sup>
df	2
Asymp. sig.	0.006

<sup>a</sup>0 is treated as a success

The test is highly significant with a p-value of 0.006. This means that there is a significant difference between the treatment responses. However, we do not know where: between treatments 1 and 2, 2 and 3, or between 1 and 3. For that purpose three separate McNemar’s tests have to be carried out.

Test statistics <sup>a</sup>	
Treatment 1 and treatment 2	
N	139
Chi-square <sup>b</sup>	4.379
Asymp. sig.	0.036

<sup>a</sup>McNemar test

<sup>b</sup>Continuity corrected

Test statistics <sup>a</sup>	
Treatment 1 and treatment 3	
N	139
Chi-square <sup>b</sup>	8.681
Asymp. sig.	0.003

<sup>a</sup>McNemar test

<sup>b</sup>Continuity corrected

Test statistics <sup>a</sup>	
Treatment 2 and treatment 3	
N	139
Chi-square <sup>b</sup>	0.681
Asymp. sig.	0.409

<sup>a</sup>McNemar test

<sup>b</sup>Continuity corrected

The above three separate McNemar's tests show that there is no difference between the treatments 2 and 3, but there are significant differences between 1 and 2, and 1 and 3. If we adjust the data for multiple testing, for example, by using  $p = 0.01$  instead of  $p = 0.05$  for rejecting the null-hypothesis, then the difference between 1 and 2 loses its significance, but the difference between treatments 1 and 3 remains statistically significant.





# Chapter 15

## Cox Regression (60 Patients)

Primary scientific question: is there a significant difference in survival between the group treated with one treatment versus the other.

1	Variable			
	2	3	4	5
1.00	1	0	65.00	0.00
1.00	1	0	66.00	0.00
2.00	1	0	73.00	0.00
2.00	1	0	54.00	0.00
2.00	1	0	46.00	0.00
2.00	1	0	37.00	0.00
2.00	1	0	54.00	0.00
2.00	1	0	66.00	0.00
2.00	1	0	44.00	0.00
3.00	0	0	62.00	0.00
4.00	1	0	57.00	0.00
5.00	1	0	43.00	0.00
6.00	1	0	85.00	0.00
6.00	1	0	46.00	0.00
7.00	1	0	76.00	0.00
9.00	1	0	76.00	0.00
9.00	1	0	65.00	0.00
11.00	1	0	54.00	0.00
12.00	1	0	34.00	0.00
14.00	1	0	45.00	0.00
16.00	1	0	56.00	1.00
17.00	1	0	67.00	1.00
18.00	1	0	86.00	1.00
30.00	1	0	75.00	1.00
30.00	1	0	65.00	1.00
30.00	1	0	54.00	1.00
30.00	1	0	46.00	1.00
30.00	1	0	54.00	1.00
30.00	1	0	75.00	1.00
30.00	1	0	56.00	1.00
30.00	1	1	56.00	1.00

(continued)

Variable (continued)				
1	2	3	4	5
30.00	1	1	53.00	1.00
30.00	1	1	34.00	1.00
30.00	1	1	35.00	1.00
30.00	1	1	37.00	1.00
30.00	1	1	65.00	1.00
30.00	1	1	45.00	1.00
30.00	1	1	66.00	1.00
30.00	1	1	55.00	1.00
30.00	1	1	88.00	1.00
29.00	1	1	67.00	1.00
29.00	1	1	56.00	1.00
29.00	1	1	54.00	1.00
28.00	0	1	57.00	1.00
28.00	1	1	57.00	1.00
28.00	1	1	76.00	1.00
27.00	1	1	67.00	1.00
26.00	1	1	66.00	1.00
24.00	1	1	56.00	1.00
23.00	1	1	66.00	1.00
22.00	1	1	84.00	1.00
22.00	0	1	56.00	1.00
21.00	1	1	46.00	1.00
20.00	1	1	45.00	1.00
19.00	1	1	76.00	1.00
19.00	1	1	65.00	1.00
18.00	1	1	45.00	1.00
17.00	1	1	76.00	1.00
16.00	1	1	56.00	1.00
16.00	1	1	45.00	1.00

Var 1 = months of follow-up (Var = variable).

Var 2 = event (lost for follow up or completed the study = 0, death = event = 1).

Var 3 = treatment modality (0 of 1).

Var 4 = age.

Var 5 = gender.

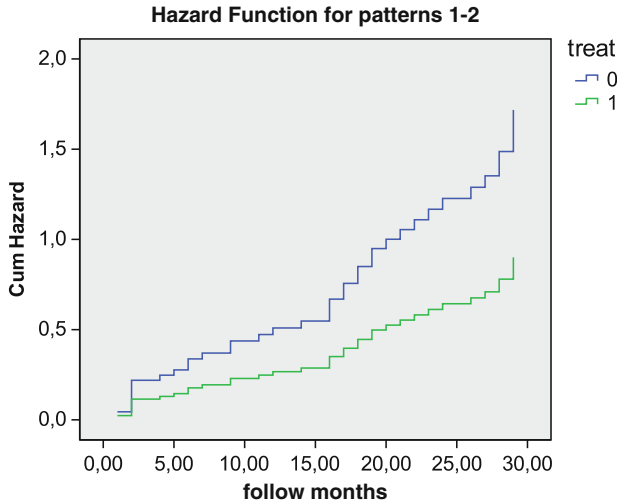
## Simple Cox Regression

Command:

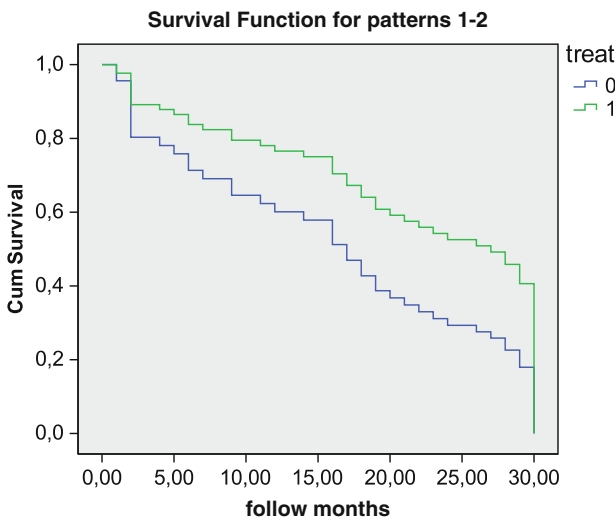
Analyze – survival – Cox regression – time: follow months – status: var 2 – define event (1) – covariates – categorical: treat → categorical variables – continue – plots – survival → var 3 → separate lines – hazard – continue – ok.

Variables in the equation						
	B	Std. Error	Wald	df	Sig.	Exp(B)
VAR00003	0.645	0.270	5.713	1	0.017	1.907

The regression coefficient, the B-value, is significantly larger than 0. The treatment modalities, treatments 1 and 2, have a significantly different effect on the chance of survival with a p-value of 0.017. The hazard ratio equals 1.907, which means that the chance of survival of one treatment is almost twice as large as that of the other treatment.



**Fig.** On the y-axis % of deaths, on the x-axis the time in months. The treatment 1 (indicated in the graph as 0) seems to cause more deaths than treatment 2 (indicated as 1)



**Fig.** On the y-axis % of survivors, on the x-axis the time (months). The treatment 1 (indicated in the graph as 0) seems to cause fewer survivors than does the treatment 2 (indicated in the graph as 1)

The interesting thing about Cox regression is that, just like with linear and logistic regression, we can use patient characteristics as additional predictors of better survival.

### Multiple Cox Regression

Before the multiple regression we first perform a simple Cox regression to find out whether gender is a significant predictor of survival.

Command:

Analyze – survival – Cox regression – time: follow months – status: var 2 – define event (1) – covariates – categorical: gender → categorical variables – continue – ok.

Variables in the equation						
	B	Std. error	Wald	df	Sig.	Exp(B)
VAR00005	-7.168	3.155	5.161	1	0.023	0.001

The above table shows that, if a simple Cox regression is performed with gender as x-variable (male/female variable 0.00005), there is, just like with treatment modality, a significant effect on survival/deaths. Gender, obviously, is also a predictor of survival. Males perform much better than females. We will now use both gender and treatment modality as predictors in order to find out whether both of them are independent determinants of the chance of surviving.

Command:

Like above, but enter both Var 00003 and Var 00005 as covariates.

Variables in the equation						
	B	Std. Error	Wald	df	Sig.	Exp(B)
VAR00005	-13.446	63.033	0.046	1	0.831	0.000
VAR00003	-0.335	0.373	0.805	1	0.369	0.716

The above multiple Cox regression with gender (variable 0.00005) and treatment modality (variable 0.00003) as predictors, appear not to produce any significant effect. Both predictors assessed simultaneously appear not to be significant factors anymore. The conclusion should be that the beneficial effect of treatment is based on confounding: if you adjust for the difference in gender the significant effect disappears. And so, the so called beneficial effect of the treatment modality is in fact caused by the fact that many more females are in one of the treatment groups.

# Chapter 16

## Cox Regression with Time-dependent Variables (60 Patients)

Scientific question: is elevated LDL-cholesterol a significant predictor of survival.

Variable					
1	2	3	4	5	6
1.00	1	0	65.00	0.00	2.00
1.00	1	0	66.00	0.00	2.00
2.00	1	0	73.00	0.00	2.00
2.00	1	0	54.00	0.00	2.00
2.00	1	0	46.00	0.00	2.00
2.00	1	0	37.00	0.00	2.00
2.00	1	0	54.00	0.00	2.00
2.00	1	0	66.00	0.00	2.00
2.00	1	0	44.00	0.00	2.00
3.00	0	0	62.00	0.00	2.00
4.00	1	0	57.00	0.00	2.00
5.00	1	0	43.00	0.00	2.00
6.00	1	0	85.00	0.00	2.00
6.00	1	0	46.00	0.00	2.00
7.00	1	0	76.00	0.00	2.00
9.00	1	0	76.00	0.00	2.00
9.00	1	0	65.00	0.00	2.00
11.00	1	0	54.00	0.00	1.00
12.00	1	0	34.00	0.00	1.00
14.00	1	0	45.00	0.00	1.00
16.00	1	0	56.00	1.00	1.00
17.00	1	0	67.00	1.00	1.00
18.00	1	0	86.00	1.00	1.00
30.00	1	0	75.00	1.00	2.00
30.00	1	0	65.00	1.00	2.00
30.00	1	0	54.00	1.00	2.00
30.00	1	0	46.00	1.00	2.00
30.00	1	0	54.00	1.00	2.00
30.00	1	0	75.00	1.00	2.00
30.00	1	0	56.00	1.00	2.00
30.00	1	1	56.00	1.00	2.00

(continued)

Variable (continued)					
1	2	3	4	5	6
30.00	1	1	53.00	1.00	2.00
30.00	1	1	34.00	1.00	2.00
30.00	1	1	35.00	1.00	2.00
30.00	1	1	37.00	1.00	2.00
30.00	1	1	65.00	1.00	2.00
30.00	1	1	45.00	1.00	2.00
30.00	1	1	66.00	1.00	2.00
30.00	1	1	55.00	1.00	2.00
30.00	1	1	88.00	1.00	2.00
29.00	1	1	67.00	1.00	1.00
29.00	1	1	56.00	1.00	1.00
29.00	1	1	54.00	1.00	1.00
28.00	0	1	57.00	1.00	1.00
28.00	1	1	57.00	1.00	1.00
28.00	1	1	76.00	1.00	1.00
27.00	1	1	67.00	1.00	1.00
26.00	1	1	66.00	1.00	1.00
24.00	1	1	56.00	1.00	1.00
23.00	1	1	66.00	1.00	1.00
22.00	1	1	84.00	1.00	1.00
22.00	0	1	56.00	1.00	1.00
21.00	1	1	46.00	1.00	1.00
20.00	1	1	45.00	1.00	1.00
19.00	1	1	76.00	1.00	1.00
19.00	1	1	65.00	1.00	1.00
18.00	1	1	45.00	1.00	1.00
17.00	1	1	76.00	1.00	1.00
16.00	1	1	56.00	1.00	1.00
16.00	1	1	45.00	1.00	1.00

Var 00001 = follow-up period (months) (Var = variable).  
 Var 00002 = event (0 or 1, event or lost for follow-up = censored).  
 Var 00003 = treatment (0 or 1, treatment 1 or 2).  
 Var 00004 = age (years).  
 Var 00005 = gender (0 or 1, male or female).  
 Var 00006 = elevated LDL-cholesterol (0 or 1, no or yes).

LDL-cholesterol is a major risk factor for cardiovascular disease.

Just like with the examples in the Chapter 11 (logistic regression) and Chapter 15 (Cox regression) a binary x-variable is used: the presence or absence of an elevated LDL-cholesterol. We want to assess whether the presence of an elevated LDL-cholesterol is a predictor of death. The hazard of death is used as the dependent variable.

## Simple Cox Regression

Command:

Analyze – survival – Cox regression – time: follow years – status: var 2 – define event (1) – Covariates – categorical: elevated LDL-cholesterol (Var 00006) → categorical variables – continue – plots – survival → hazard – continue – ok.

Variables in the equation						
	B	Std. error	Wald	df	Sig.	Exp(B)
VAR00006	-0.482	0.307	2.462	1	0.117	0.618

Var 00006 is a binary variable for LDL-cholesterol. It is not a significant predictor of survival with a p-value and a hazard ratio of only 0.054 and 0.524 respectively, as demonstrated above by a simple Cox regression with event as outcome variable and frailty as predictor. The investigators believe that the presence of LDL-cholesterol must be a determinant of survival. And if we look at the data, we will observe that something very special is going on: in the first decade virtually no one with elevated LDL-cholesterol dies. In the second decade virtually everyone with an elevated LDL-cholesterol does: LDL-cholesterol seems to be particularly a killer in the second decade. Then, in the third decade other reasons for dying seem to have taken over. In order to assess whether elevated LDL-cholesterol adjusted for time has a significant effect on survival, a time-dependent Cox regression will be performed. For that purpose the time-dependent covariate is defined as a function of both the variable time (called “T\_” in SPSS) and the LDL-cholesterol variable, while using the product of the two. This product is applied as “time-dependent predictor of survival, and a usual Cox model is, subsequently, performed (Cov = covariate).

## Cox Regression with Time-dependent Variables

Command:

Analyze – survival – Cox w/time-dependent cov – compute time-dependent cov – Time (T\_) → in box expression for T\_Cov – add the sign \* – add the LDL-cholesterol variable – model – time: follow months – status: var 00002 – ?: define event:1 – continue – T\_Cov → in box covariates – ok.

Variables in the equation						
	B	Std. error	Wald	df	Sig.	Exp(B)
T_COV_	-0.131	0.033	15.904	1	0.000	0.877

The above results table of the “Cox regression with time-dependent variables” shows that the presence of an elevated LDL-cholesterol adjusted for differences in time is a highly significant predictor of survival.

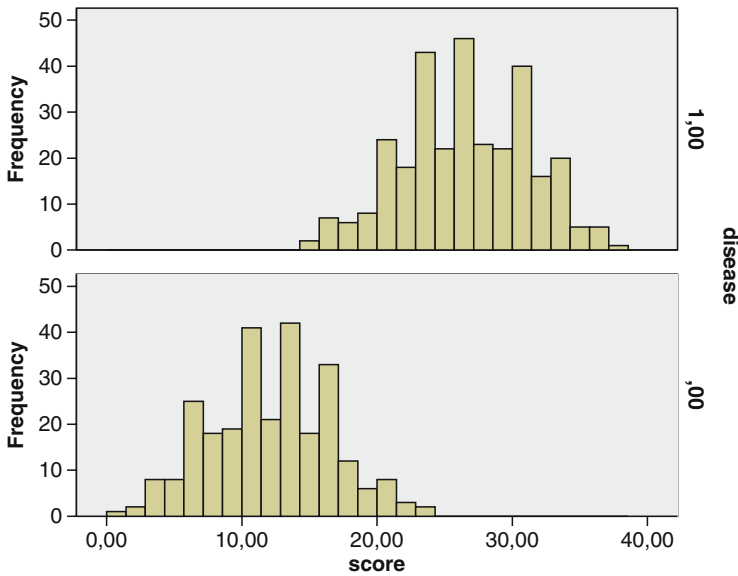






First we try and make a graph of the data.

Command:



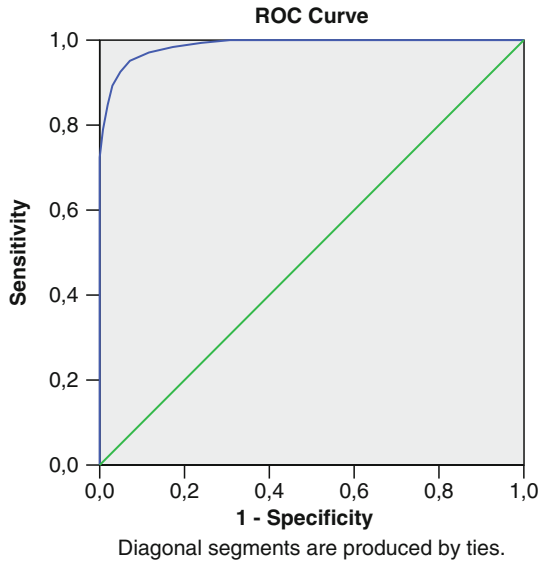
Analyze – legacy dialogs – graphs – histogram – variable: score – rows: disease – ok.

The above histograms summarize the data. The upper graph shows the frequencies of various scores of all patients with vascular disease as confirmed by angiograms, the lower graph of the patients without. The scores of the diseased patients are generally much larger, but there is also a considerable overlap. The overlap can be expressed by sensitivity (number of true positive/number of false and true positive patients) and specificity (number of true negative patients/number of false and true negative patients). The magnitude of the sensitivity and specificity depends on the cutoff level used for defining patients positive or negative. sensitivities and specificities continually change as we move the cutoff level along the x-axis. A Roc curve summarizes all sensitivities and specificities obtained by this action. With help of the Roc curve the best cutoff for optimal diagnostic accuracy of the test is found.

### Validating the Qualitative Diagnostic Test

Command:

Graphs – ROC curve – test variable score – state variable: disease – value of state: variable 1 – mark: ROC curve – mark: with diagnostic reference line – mark: coordinate points of ROC curve – ok.



The best cutoff value of the sensitivity and 1-specificity is the place on the curve with the shortest distance to the top of y-axis where both sensitivity and 1-specificity equal 1 (100%). The place is found by adding up sensitivities and specificities as summarized in the following table.

Coordinates of the curve		
Test result variable(s): score		
Positive if greater than or equal to <sup>a</sup>	Sensitivity	1-specificity
0.0000	1.000	1.000
1.5000	1.000	0.996
2.5000	1.000	0.989
3.5000	1.000	0.978
4.5000	1.000	0.959
5.5000	1.000	0.929
6.5000	1.000	0.884
7.5000	1.000	0.835
8.5000	1.000	0.768
9.5000	1.000	0.697
10.5000	1.000	0.622
11.5000	1.000	0.543
12.5000	1.000	0.464
13.5000	1.000	0.382
14.5000	1.000	0.307
15.5000	0.994	0.240
16.5000	0.984	0.172
17.5000	0.971	0.116
18.5000	0.951	0.071
19.5000	0.925	0.049

(continued)

Coordinates of the curve (continued)

20.5000	0.893	0.030
21.5000	0.847	0.019
22.5000	0.789	0.007
23.5000	0.724	0.000
24.5000	0.649	0.000
25.5000	0.578	0.000
26.5000	0.500	0.000
27.5000	0.429	0.000
28.5000	0.354	0.000
29.5000	0.282	0.000
30.5000	0.214	0.000
31.5000	0.153	0.000
32.5000	0.101	0.000
33.5000	0.062	0.000
34.5000	0.036	0.000
35.5000	0.019	0.000
36.5000	0.010	0.000
37.5000	0.003	0.000
39.0000	0.000	0.000

The test result variable(s): score has at least one tie between the positive actual state group and the negative actual state group

<sup>a</sup>The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values

The best cutoff value of the sensitivity and 1-specificity is the place on the curve with the shortest distance to the top of y-axis where both sensitivity and 1-specificity equal 1 (100%). The place is found by adding up sensitivities and specificities as summarized in the underneath table.

Sensitivity	1-specificity	sensitivity – (1-specificity) (= sensitivity + specificity-1)
0.971	0.116	0.855
0.951	0.071	0.880
0.925	0.049	0.876

At a sensitivity of 0.951 and a “1-specificity” (= false positives) of 0.071 the best add-up sum is found (1.880). Looking back at the first column of the table from the previous page the cutoff score >18.5 is the best cutoff, which means a score of 19 produces the fewest false positive and fewest false negative tests.

## Chapter 18

# Validating Quantitative Diagnostic Tests (17 Patients)

Primary scientific question: is angiographic volume an accurate method for demonstrating the real cardiac volume.

Variable	
1	2
494.00	512.00
395.00	430.00
516.00	520.00
434.00	428.00
476.00	500.00
557.00	600.00
413.00	364.00
442.00	380.00
650.00	658.00
433.00	445.00
417.00	432.00
656.00	626.00
267.00	260.00
478.00	477.00
178.00	259.00
423.00	350.00
427.00	451.00

Var 00001 = cast cardiac volume (liters) (Var = variable)

Var 00002 = angiographic cardiac volume (liters)

## Validating Quantitative Diagnostic Test

Command:

Analyze – regression – linear – dependent = cast cardiac volume – independent = angiographic cardiac volume – ok.

		Coefficients <sup>a</sup>				
		Unstandardized coefficients		Standardized coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	39.340	38.704		1.016	0.326
	VAR00001	0.917	0.083	0.943	11.004	0.000

<sup>a</sup>Dependent variable: VAR00002

Four tables are given, but we use the bottom table as shown above.

B = regression coefficient =  $0.917 \pm 0.083$  (std. error).

A = intercept (otherwise called  $B_0$  or Constant) =  $39.340 \pm 38.704$  (std. error).

The 95% confidence intervals of A and B should not be different from respectively 0.000 and 1.000. This can be confirmed, because they are respectively.

Between  $0.917 \pm 1.96 \times 0.0813$ , and thus between 0.751 and 1.083.

Between  $39.340 \pm 1.96 \times 38.704$ , and thus between  $-38.068$  and 116.748.

This diagnostic test is thus, accurate.



### Analysis: Calculate Cohen’s Kappa

Command:

Analyze – descriptive statistics – crosstabs – variable 1 click into rows – variable 2 click into columns – statistics – mark: kappa – continue – crosstabs: mark: cells – cell display: mark observed (under counts) and total (under percentages) – continue – ok.

Symmetric measures				
	Value	Asymp. std. error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. sig.
Measure of agreement Kappa	0.400	0.167	2.196	0.028
No. of valid cases	0.30			

<sup>a</sup>Not assuming the null hypothesis

<sup>b</sup>Using the asymptomatic standard error assuming the null hypothesis

The above table shows that the kappa-value equals 0.400. A kappa-value of 0 means poor reproducibility or agreement, a kappa-value of 1 means excellent. This result of 0.400 is moderate. This result is significantly different from an agreement of 0 at  $p = 0.028$ .



## Chapter 20

# Reliability Assessment of Quantitative Diagnostic Tests (17 Patients)

Primary scientific question: is the underneath quantitative diagnostic test adequately reproducible.

Variables	
1	2
10.000	10.000
9.00	10.000
7.00	6.00
5.00	6.00
3.00	7.00
8.00	8.00
7.00	7.00
8.00	7.00
7.00	8.00
8.00	8.00
7.00	9.00
10.000	11.00

Var 1 = quality of life score at first assessment (Var = variable).

Var 2 = quality of life score art second assessment.

## Intraclass Correlation for Reliability Assessment of Diagnostic Battery

Command:

Analyze – Scale – Reliability Analysis – Var 00001 and Var 00002 click into items box – Statistics – Mark: Intraclass correlation coefficient – Model: Two-way Mixed – Type: Consistency – Test value: 0 – Continue – ok.

Intraclass correlation coefficient							
		95% confidence interval		F test with true value 0			
	Intraclass correlation <sup>a</sup>	Lower bound	Upper bound	Value	df1	df2	Sig
Single measures	0.712 <sup>b</sup>	0.263	0.908	5.952	11	11	0.003
Average measures	0.832 <sup>c</sup>	0.416	0.952	5.952	11	11	0.003

Two-way mixed effects model where people effects are random and measures effects are fixed.

<sup>a</sup>Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance

<sup>b</sup>The estimator is the same, whether the interaction effect is present or not

<sup>c</sup>This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise

The above table shows that the intraclass correlation (= SS between subjects/(SS between subjects + SS within subjects), SS = sum of squares) equals 0.832 (= 83%) if interaction is not taken into account, and 0.712 (=71%) if interaction is accounted. An intraclass correlation of 0 means that the reproducibility/agreement between the two assessments in the same subject is 0, 1 indicates 100% reproducibility/agreement. An agreement of 40% is moderate and of 80% is excellent. In the above example there is, thus, a very good agreement with a p-value much smaller than 0.05, namely 0.003. The agreement is, thus, significantly better than an agreement of 0%.

# Chapter 21

## Final Remarks

Note:

1. Rank tests with ties: tie is rank number with multiple values.
2. Continuity correction for  $\chi^2$  test: if discrete variables are approximated by the normal distribution (= continuous), then a better approximation is obtained by adding +1/2 (or subtracting) to the probabilities, for example:  $\chi^2 = (O - E)^2/E$  replace with  $= (|O - E| - 0.5)^2/E$ .
3. Unrequested alternative tests (contrast tests) are often included gratuitously in the analysis for example with the chisquare test).
4. Logistic regression: a lot of tables prior to the actual testing is produced. This is predominantly meant for assessing the goodness of fit of the so-called loglinear model. Look for more information in the textbook “Statistics Applied to Clinical Trials, Springer Dordrecht, 2009, by the same authors. If the data do not adequately fit the model, then the final analysis is often not executed anymore. So, you do not have to worry too much about this issue.
5. The four non-parametric tests are probably the most valuable tests for the analysis of data from clinical trials (Wilcoxon, Mann–Whitney, Friedman, Kruskal–Wallis), and they are particularly safe for those who are not fond on statistics but rather on solid results.



# Index

## A

- Accuracy
  - of qualitative diagnostic tests, 64
  - of quantitative diagnostic tests, 37
- Agreement, 70, 72
- Analysis of variance (ANOVA), 1, 16–18, 21–23, 29–31, 34

## B

- Binary data, 1, 3–5, 35–37, 43–45, 51–53
- Binary logistic regression, 41, 42
- Bonferroni adjustment, 23
- B-value, 17, 19, 57

## C

- Chi-square test, 37, 41, 44–45
- Chi-square test for trends, 45
- Cochran's Q test 61, 51–53
- Cohen, 70
- Columns, 2, 27, 37, 44, 66, 70
- Continuity correction, 37, 73
- Continuous data, 1, 3, 7–9, 11–13, 33–34
- Contrast tests, 37, 73
- Correlation coefficient, 16, 71, 72
- Cox regression, 1, 55–61
- Cox regression with time-dependent variables, 59–61
- Crosstabs, 5–37, 44, 70
- Cutoff value for diagnostic test, 65, 66

## D

- Degrees of freedom, 4, 22, 26
- Diagnostic tests, 1, 63–70

## F

- Friedman, 1, 21–23, 30, 73

## G

- Gaussian data, 9, 13, 17, 22
- Greenhouse method, 22

## H

- Hazard function, 57
- Hazard ratio, 57, 61
- Histogram, 64
- Huynh method, 22

## I

- Interaction variable, 42
- Intercept, 22, 27, 68
- Intraclass correlation, 71–72

## K

- Kappa, 70
- Kruskall-Wallis, 1, 29–31, 73

## L

- Linear-by-linear association, 37, 44, 45
- Linear regression, 15–19, 34, 41, 67
- Logistic regression, 39–42, 45, 57, 60, 73
- Loglinear model, 73
- Lower-bound method, 22

## M

- Mann-Whitney, 11–13, 31, 73
- Mauchly's test, 22
- McNemar's chi-square test, 47–49, 52, 53

- Measure of agreement, 70  
Mixed model for repeated measures, 21–23, 26  
Mixed models, 25–27  
Multiple Cox regression, 58  
Multiple linear regression, 18–19  
Multiple logistic regression, 42  
Multiple paired binary data, 1, 51–53  
Multiple testing, 22, 23, 30, 53
- N**  
Non-Gaussian data, 23, 30–31  
Non-parametric tests, 73  
Null-hypothesis, 22, 30, 53
- O**  
Odds of responding, 44, 45  
Odds ratio, 41, 42  
One-sample t-test, 3–4  
One-way-ANOVA, 29–31
- P**  
Paired analysis of variance, 1  
Paired binary data, 1, 51–53  
Paired continuous data, 7–9  
Paired t-test, 8–9, 22  
Parametric tests, 1  
Pearson chi-square, 37, 44, 45  
Positive correlations, 26, 52  
Proportions, 5, 22  
p-values, 9, 13, 17, 19, 22, 23, 30, 31, 34, 37, 41, 42, 45, 49, 52, 57, 61, 72
- R**  
Rank tests with ties, 73  
Regression coefficient, 17, 19, 41, 57, 68  
Regression equation, 19  
Reliability assessment  
    of qualitative test, 69–70  
    of quantitative test, 71–72  
Repeated-measures-ANOVA, 21–23, 26  
Reproducibility, 70, 72  
ROC curves, 64, 65
- Rows, 2, 37, 44, 64, 70  
R-square value, 12, 17–19  
R-value, 16
- S**  
Sensitivity, 34, 63–66  
Simple Cox regression, 56–58, 61  
Simple linear regression, 16–18, 34, 41  
Simple logistic regression, 41  
Specificity, 63–66  
SS between subjects, 72  
SS within subjects, 72  
Summary estimates, 26
- T**  
Three dimensional charts, 37, 44, 49  
Ties, 13, 65, 73  
Time-dependent Cox regression, 61  
Trend tests  
    of binary data, 43–45  
    of continuous data, 33–34  
t-value, 4, 17, 26
- U**  
“Unpaired” analysis of variance, 1  
Unpaired binary data, 1, 35–37  
Unpaired continuous data, 11–13  
Unpaired t-test, 11–13, 15, 26
- V**  
Validating qualitative diagnostic tests, 63–66  
Validating quantitative diagnostic tests, 67–68
- W**  
Wilcoxon, 7–9  
Wilcoxon signed ranks test, 9
- Z**  
z-test, 3–5  
z-value, 5