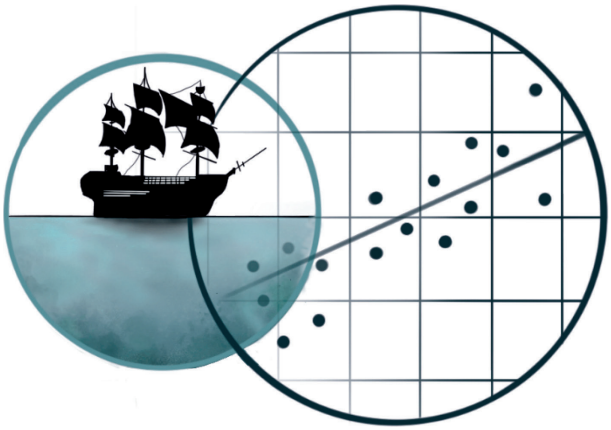
VINCE PENDERS

PIRATES, PEACHES AND P-VALUES

STYLISH STATISTICS!



PARRRT 2

All rights reserved

Copyright 2019 Piraten, perziken en p-waarden, Maastricht
www.pppwaarden.nl

Production and distribution: Mosae Verbo, Maastricht
www.boekenplan.nl

Cover and illustrations: Chelsy Penders
Formatting: Chelsy and Vince Penders
Editing: Vince Penders and Sophia von Stockert

ISBN 978 90 8666 485 6
NUR 123

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of the publisher, nor be otherwise circulated in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

TABLE OF CONTENTS

THOSE WHO CREATED SALUTE YOU **7**
SAILORS' TALES **8**
WHY WOULD YOU USE THIS BOOK? **10**
HOW TO USE THIS BOOK? **12**

OVERVIEWS AND CORE CONCEPTS **13**

DESCRIBING DATA **29**
Chapter 20 Advanced Data Explorations 30

ANALYSIS OF VARIANCE **49**
Chapter 21 Bonferroni and Contrast Analyses 50
Chapter 22 Two-way ANOVA 74
Chapter 23 ANCOVA 109
Chapter 24 Within-subjects ANOVA 130
Chapter 25 Two-way Within-subjects ANOVA 160
Chapter 26 Split-plot ANOVA 176
Chapter 27 ANCOVA for Repeated Measures 203
Chapter 28 MANOVA and Discriminant Analysis 220

CATEGORIES AND CROSS TABULATIONS **247**
Chapter 29 Contingency Table Analysis 248

REGRESSION **277**
Chapter 30 Dichotomous and Dummy Variables 278
Chapter 31 Multiple Regression 295
31-A Main Effects 297
31-B Interaction and Simple Slopes 322
31-C Mediation 337
Chapter 32 Logistic Regression 343
32-A Main Effects 344
32-B Interaction Effects 356
Chapter 33 Multi-level Regression 369
33-A Nested Designs 370
33-B Repeated Measures 402

METHODOLOGY **433**
Chapter 34 Power Analysis 434

PSYCHOMETRICS **461**
Chapter 35 Factor Analysis 462
Chapter 36 Reliability 492
Chapter 37 Validity and Diagnostics 518

BUILDING BRIDGES **531**
Chapter 38 Contrasts and ANOVA 532
Chapter 39 ANOVA and Regression (Parrrt II) 534
Chapter 40 Contingency Table Analysis and Logistic Regression 547

APPENDIX **551**
Bibliography 551
Action Plan for Analyses 555
Choosing a test 568
Statistical tables 570
Index 573

THOSE WHO CREATED SALUTE YOU

Welcome, dear reader! A brand-new edition of this handbook is sparkling in your hands, and boy, are we proud of the result. We, the makers, would like to introduce ourselves briefly. Is that all right?

The guy on the left is **VINCE PENDERS**… just call me **Vincenzo**. The brain behind this teaching method – that's me. Creative to the bone, a cheerful perfectionist. Previously employed at Maastricht University, I work as an independent teacher nowadays, and tutor roughly 300 students per year while constantly improving my book. On those rare occasions when I'm not working or seeing friends, you will find me cooking, playing a Nintendo game, or writing fiction with epic music in my ears. That's right: novels are my other passion. The upcoming one, *Een beer en een boerka*[1], tells the tale of a man who believed that he was something small and fluffy. Like my debut novel *Zwaluwhart*, it will be published by Macc in 2020.

My sister **CHELSY PENDERS** (in front) sings, dances and works as an interior decorator. Where my artistic voice is the written word, hers is the hand-made drawing in every shape and style. She will take on any project that appeals to her visual talents. Chelsy's versatility is obvious from the many illustrations that populate this book, as well as the slick lay-out with its gorgeous colour schemes (please excuse my lack of pride for her work). Her Maine Coon cats love getting cuddly when she's drawing at home, so if you spot a cat paw somewhere, make sure to let us know.

And this is **SOPHIA VON STOCKERT**, a former student of mine and a good friend nowadays. She is passionate about research, statistics and writing. A dedicated statistics tutor in the past, she inspired students to get into all the cool stuff that one can do with stats and methods. This is why I brought her on board as an editor of *Pirates, Peaches and P-values*. Sophia holds a Research Master's degree in Cognitive and Clinical Neuroscience, and currently works as a data analyst in Berlin. In her free time she goes bouldering and reviews and translates books and manuscripts.

What remains are a few honourable mentions as ways to express my gratitude. The first one goes out to Ward Schoonbrood, my good friend who manages the PPP website and is always happy to give me careful advice. The second one is for René Bouman, the publisher who aided us in realising this book. Finally, a big hug to all my students who inspire me every day. Helping you is an honour. You convinced me to write this book and kept my passion burning to improve it further with this new edition. People often thank me for the work I do, but there's only one proper answer. "No… thank *you*!"

[1] Yeah, sorry, it's currently in Dutch. The title translates to *A Bear and a Burqa*.

Sailors' tales

I would definitely recommend this book to people who have a hard time getting statistics. PPP allows you to understand statistics in entertaining and concrete problems. Vince never fails to make me chuckle and his sister Chelsy brings colour to math!
– Hana Vacková

I bought Vince's book at the end of my first year and immediately regretted that I was that late.
– Ekin Tünçok

This book helped me greatly. It perfectly combines creativity and humor with statistics, making it a genuinely enjoyable read. More importantly, it helps you understand and remember everything much better. You just won't forget examples that include ninjas, socially anxious vampires and orcs on a diet!
– Elin Hammarberg

The perfect guide for all statistics courses.
– Anna Bergauer

It's easy to be the captain in calm waters. But when the storms of statistics take over, would you rather have many dark nights... or one of Vince's spinach delights? Climb aboard the PPP ship and find yourself sailing along your own learning curve. From variable winds to confident cruising; this is the treasure map you've been looking for, and an adventure I can wholeheartedly recommend.
– JCCSS

Teaching is an art, and this book taught me that having fun is the best way to learn!
– Micol Yannuzzi

I grew from a peach to a pirate and conquered the p-values – a journey I would not have succeeded in if it was not for our Captain. Working with this book literally saved me from drowning in the stormy seas of statistics. The excitingly intricate examples and thorough, easy-to-grasp explanations changed my whole perspective on the subject. Statistics is what one makes of it and Vince made the best of it!
– Leonie Ambraß

Illustrative examples; step-by-step guidance; utterly relevant overviews, tables, figures and diagrams with a clear and catchy format. Vince engages with the reader in an empathic and encouraging way as he addresses the various questions that may arise in the reader's mind. This is a lifelong reference book accessible to everyone interested in science and statistics. Learning with confidence and curiosity is guaranteed!
– Ege Idil Bor

This book is just Vinceceptional. Concepts are explained in a very clear and simple way. With no background in statistics, I was able to catch up for a Research Master's course in just a few weeks and developed a passion for statistics. Thanks, Vince!
– Robert Schreiber

The closest you can get to purchasing a 'statistics bible'. It touches upon all the necessary materials, but in a unique and innovative way.
– Fredy Enders

WHY WOULD YOU USE THIS BOOK?

DEAR STUDENT,

You probably came here in pursuit of the rumours: rumours about a handbook that does things differently. A teaching method that understands you, that takes you seriously, and that can truly improve your insight. Statistics is a subject you've never asked for; you study Psychology, Medicine or Social Sciences, and are not an expert at mathematics. But since statistics is so important in your field, you can't get around it. Will it ever become more exciting and tangible than a game of Xs and Ys?
As the responses show: yes. Anyone – and I'm not exaggerating – can see through statistics this way. Before you lies my approach of clear, crazy and complete explanations, translated to paper. You will test cupcakes, experiment on civilians in a zombie apocalypse, observe witches and wizards and much more. It will make you smarter than you had deemed possible until now.
In short: try it, and be convinced. Help build the revolution, along with the 1600 students that came before you. Still think you're a peach in statistics? You're about to become a pirate!

Greetings,
Vincenzo

DEAR TEACHER,

You may open this handbook with a touch of scepticism. Why did such a whippersnapper decide to launch his own teaching method? Should he not have left that to seasoned professors, whose years of teaching and research have allowed them to grow into experts? Really, what has gotten into him, to tempt students into putting aside the recommended literature and change to his unsolicited alternative?
The claim that I possess less expertise than a scientist with a PhD in statistics needs no long debate: you are undoubtedly right. I have a lot to offer, but I know my limits. Perhaps part of my strength arises from this fact. I can empathise with the student, who sometimes has to start from scratch, and often has no sense of mathematics or has yet to develop it. You can guess that I chose unorthodox examples and a light-hearted tone for this very reason. Chances are you will think that both are simply fineries, or even – in the most unfavourable case – that they distract the student from the things he or she should be learning. My view is different. Allow me to provide an explanation.
First, I suspect that *structure* and *a clear overview* are two of the key pillars a student needs. Apart from my examples and style, I have done my best to offer these. Most people experience the organisation of new knowledge as exceptionally challenging, especially if the field does not suit them initially. By conducting some preliminary work, I prevent a lot of frustration. Each chapter begins with sharp tabulated summaries of the theory that is discussed, which show a clear continuity throughout the method: research designs, mathematical formulas, assumptions of statistical tests, and action plans for an analysis are mentioned point by point. The general *Action Plan for Analyses*, found in the Appendix module of Parrrt 2, offers further support in developing a bird's-eye view, and can be used at later stages as well by students and researchers to properly steer their statistics. You may hold the opinion that I take too much work out of the students' own hands. In my own experience, however, they learn

most effectively if they are presented with this structure at least once. In contexts that allow for interaction, such as a lecture or lesson, a middle road can be walked which is the optimum in my view: together with the teacher, the students will then fill in the summaries themselves. I also advise readers of this handbook to do the same thing in their own notes (see *How to use this book?*).

Second, I am convinced that a teaching method for statistics should be both accessible and complete. Accessible in the sense that anyone should understand it; complete in the sense that it skips as few steps as possible. Roughly, statistical education knows two excesses. One excess is the 'just take this formula and fill it in' approach; you spare the student the details, but he or she still has no idea what it is actually about and remains deprived of true insights. At the other end we find formal mathematical definitions, where abstract language, matrices, proofs and subscripts build a heavily fortified castle without a visitor's entrance – and never forget: the student is a visitor in the statistical universe, an immigrant who still has to find his way. Neither of these extremes please me. This book is a roadmap which seeks to stop by a great many locations, using paths without pitfalls, and with tips for the connoisseur. When a formula presents itself, I will tell you how it was constructed; if a statistical test makes a certain assumption, I will explain why. Should a problem truly carry too far for a regular course in statistics, I will offer the solution in footnotes and bonus sections for the reader who wants to go the extra mile. Many students (especially from university) find it hideous to just assume things; they prefer to go through fire and water and root their knowledge into solid ground – so afterward, they will retain that knowledge better, guaranteed.

Third, all my examples are parodies on existing research. When designing a new one, I always ask myself the same question. Will this parody help or will it distract? Is it a bizarre attempt to be interesting, or can a student really learn from it? Of course some will raise their eyebrows when asked to experiment on mermaids, observe extraterrestrial seagulls, and hand out questionnaires to grumpy orcs. Why does this method not cut to the chase? But soon these same students discover how the data come to life inside their heads. Relationships and effects take shape in their imagination, and at this point, I give those relationships and effects a statistical face. Students start to see what this mathematical gibberish is supposed to say, what significance and confounding and interaction mean. The irony? They do not learn in spite of, but *thanks to* the cheerful examples, which to them are infinitely more tangible than a variable X and a variable Y. And when he or she later tries to remember how multiple regression went again, which memory might surface faster? 'Oh, that thing with the Xs!'… or rather: 'Oh, that thing with the pirates!'?

In short: *Pirates, Peaches and P-values* wants to build bridges. A bridge between exhaustiveness and accessibility, one between substance and mathematics, and one between comedy and knowledge. And let us not forget: a bridge between teachers. As I frankly admitted earlier, your expertise in statistics is larger than mine. Thus, this book has been designated for mutual cooperation from its first paragraph. I think that you and I can complement each other from our own specialty, to finally rob the subject of statistics from its persistent status of 'difficult' and 'boring'. Should you see conditions arise on which this teaching method would be a valuable addition to your curriculum, please do not hesitate and contact us via www.pppwaarden.nl. I look forward to our first conversation.

Warm regards,
Vince Penders

HOW TO USE THIS BOOK?

Pirates, Peaches and P-values is able to function as an independent course in statistics. In practice, most readers will probably use it to support them during a college or university programme. That's why, from this edition onward, we've split the book into **two volumes**:

- **Parrrt 1** covers the fundamentals. Describing samples, probability theory, hypothesis tests, confidence intervals… You're likely to find this volume useful if you're a college student, or in your first year of university.
- **Parrrt 2** opens up the full arsenal. Statistical models with more than two variables, repeated measures designs, psychometrics… Head for these waters in your second year of university and beyond.

Of course this is just a general guideline. Individual curricula may discuss some topics from Parrrt 2 in college, or come back to stuff from Parrrt 1 during the second university year. Researchers (and perfectionistic students) may want to possess both books so that they can read up on the fine details. Note that those who only have Parrrt 2 need not fret: many key concepts from Parrrt 1 are summarised in the *Overviews and Core Concepts* section! Make sure to use it for a smooth experience. ☺

Apart from theoretical explanations, most chapters in *PPP* also offer **exercises**. I have designed them to not only test your knowledge, but add deeper layers to it as well. So practice! Well-made exercises are educational and add a lot to your hands-on experience.

If you find that something crucial is **missing** from this handbook, let us know via www.pppwaarden.nl! We look forward to making the next version even more complete.

DO YOU STUDY STATISTICS INDEPENDENTLY?

Just start at chapter 1 or at the topic you'd like to learn something about. See the next heading for following considerations.

ARE YOU TAKING A STATISTICS COURSE AT YOUR SCHOOL, COLLEGE OR UNIVERSITY?

Go to the table of contents and look for the subject you need to study. If you can't find it, consult the index at the end of the book. Topics with a green marking are also featured in the *Overviews and Core Concepts* section. If all this yields no results, you may visit my contact page on the website (www.pppwaarden.nl) and send me an e-mail, to ask whether I discuss your subject material somewhere in this book.

DO YOU NEED STATISTICS FOR YOUR OWN RESEARCH?

You are probably familiar with the topics I discuss in this book. You've forgotten a number of things though, or you would like to verify some details. In that case it's important for you to easily find what you're looking for. Study the **Action Plan for Analyses** in the Appendix in any case – it tells you not only what to do, but also how to report the results in APA style. As soon as you've chosen the statistical analysis that fits your study, you can proceed to the corresponding chapter. Between the lines you will find instructions to command SPSS as well. Should you need to refresh your knowledge about a specific concept on the side, the section *Overviews and Core Concepts* offers assistance. I hope your supervisor will be impressed by your statistics arsenal.

OVERVIEWS AND CORE CONCEPTS

The content offered in this handbook is comprised of many concepts. Some of them consistently return. This overview discusses them one by one, and in so doing it wants to be a base you can always fall back on. Don't use it to learn new things, but return to it whenever you've forgotten something familiar. When a concept is marked green in the chapters, it's meant to redirect you to this prelude.

ASSUMPTIONS

Introduced in chapter 7 (Parrrt 1)

All statistical models and hypothesis tests make a number of so-called assumptions. Those are things that the models and tests take for granted in their calculations, such that the outcomes will only make sense if those assumptions were correct.
¿Que? many readers will think at this point, so let me make this more tangible. Take a common kitchen scale. I can use this instrument to weigh a sack of sweet potatoes (or bell peppers, if those are more your thing). It will tell me that my sweet potatoes weigh 1,2 kilograms (or 2,6 pounds). But the kitchen scale assumes that I'm using it in a certain location… That's right: on the planet Earth.[2] I could travel to the moon and use my scale there; you can bet that a number of kilograms will appear on the display. However, this number will be completely *wrong*, because the scale has been configured for our planet's gravity level. The point is that my instrument will always give me a result, but does that result make sense? Only if the instrument's assumptions have been met.
It goes exactly the same for statistical tests. They rely on certain assumptions, and we can only trust the test results under these conditions. But! There are exceptions that can protect the tests from falling apart. Exceptions that make the tests **robust** against certain violations, as statisticians say. Mind the colour codes in the chapters: you cannot bypass assumptions printed in red, but you can bypass those printed in blue.

BIASED ESTIMATOR
>>> Estimate

BONFERRONI CORRECTION

Introduced in chapter 10 (Parrrt 1)

Sometimes we use multiple tests in succession to assess relationships between variables. These situations include:

- Pairwise comparisons: comparing more than two groups on the same dependent variable. In chapter 10, we studied three treatments to make mermaids more satisfied with their bodies. Eventually we compared all three conditions one-on-one, using three separate t-tests.
- Planned contrasts: in chapter 21, we studied four treatments to help vampires make more human victims. Several clusters of groups were contrasted, and two of these contrasts were not orthogonal (see elsewhere).

[2] No, not in the kitchen. Smartass.

- Simple effects analyses: in chapter 22, we discovered that the cultural backgrounds of gamblers affected the risks they took (i.e. how much money they bet). However, this effect was stronger when the gamblers were drunk. So after finding this significant interaction effect (see elsewhere), it makes sense to look at the background effect for sober and drunken gamblers separately, using two one-way ANOVAs or two t-tests.
- Studies with multiple dependent variables: in chapter 28, we found out that the way in which Egyptian women communicated with the god Anubis affected their perception of him. Namely, when they talked face to face rather than in writing, he was perceived as more powerful (dependent variable one) but less sexy (dependent variable two). Among other things, we could conduct two separate univariate ANOVAs – one for each dependent – to see whether those differences in perception were significant.

However, a problem sticks to this approach. Each time you perform a statistical test, you stand a chance of making a type I error. So by conducting multiple tests, you make the probability of at least one type I error larger than normal. If you run three tests, it can sometimes grow to almost 15% (provided that $\alpha = 0{,}05$ for each test). We call this probability the **family-wise error rate**: the extent to which an error may be made in a 'family' of tests.

The Bonferroni correction is a basic method to reduce the family-wise error rate back to 5%. It consists of dividing the significance level by the number of tests:

$$\boldsymbol{p \quad vs. \quad \frac{\alpha}{\#tests}}$$

The significance level is thus lowered. P-values must now be compared to the adapted significance level. Only if they're lower than this lowered boundary, we declare the test significant. This makes it harder to reject the null hypotheses, and less likely that we reject a true one (type I error).

You can also apply the Bonferroni correction by leaving the significance level untouched, and multiplying each p-value by the number of tests instead:

$$\boldsymbol{p * \#tests \quad vs. \quad \alpha}$$

This is effectively the same as the first method (the left- and right-hand side of the equation have just been multiplied by the number of tests). It therefore yields the exact same results. I've grown fond of this second approach, but it's a matter of personal taste.

Some final notes:

- For pairwise comparisons, there exist many other correction methods. I'm not a huge fan of the lion's share, to be honest (they're abstract and most don't give you much more power in exchange for the trouble). Check out chapter 21, section 21.2.III, for some suggestions.
- Adjusted Bonferroni correction is – in my opinion – a neat advanced variation on basic Bonferroni. I always recommend it to my senior students. See chapter 21, section 21.2.II.

CENTRAL LIMIT THEOREM
>>> Sampling distribution

CENTRING
>>> Data transformation

COHEN'S *d*
>>> Effect size

CONFIDENCE INTERVAL

Introduced in chapter 6 (Parrrt 1)

To get an impression of a population parameter, such as a mean, a popular option is to create a confidence interval. This is because sample estimates will never be perfectly accurate. For instance, in chapter 6, a sample of aliens from the planet Coronyyyr was measured to find out how many arms the average Coronyyyrian has. The mean number of arms in the sample was 4,27. The population mean, however, could be a bit different. This is why we made a 95% confidence interval: we put a margin of error around the sample mean, to get an impression of where the population mean should lie between. The interval was [3,11; 5,47].

The calculation of confidence intervals has a consistent structure, no matter what kind of parameter they're about:

$$\textit{sample estimate} \pm \textit{critical value} * \textit{standard error}$$

The critical value is a z- or t-value, which indicates how many standard errors we use for the margin of error. (For a brief recap of the standard error, see *Sampling distribution*.) A higher critical value results in a higher confidence level (say, not 95% but 99%): the confidence interval gets broader, so we can be more confident that it embraces the population parameter of interest. However, a broader interval is also less precise. To narrow it down again without sacrificing confidence, you can try to draw a bigger sample – which reduces the standard error.

Confidence intervals can also be used as two-sided hypothesis tests. Look back at the interval for the aliens on Coronyyyr. If the population mean lies between 3,11 and 5,47 indeed, we can say something about many null hypotheses. 2 arms, for instance, is *not* a plausible population mean; the null hypothesis that $\mu = 2$ must therefore be rejected. We would also reject $H_0: \mu = 3$. However, 4 or 5 are plausible values for the population mean, so we cannot reject the null hypotheses that $\mu = 4$ or $\mu = 5$. Essentially, the confidence interval tells you about all the null hypotheses that you would not reject (at a significance level of 1 minus the chosen confidence level: $\alpha = 1 - C$).

CONFOUNDING

Introduced in chapters 22, 29 and 31

This is a measuring problem. It occurs when we fail to observe the 'real' effect that a certain independent variable has, because a different independent variable messes with our observation. For instance, in chapter 29, Granger discovered that pure-blooded wizards (who had been born in an all-wizard family) were more likely to convert to the Dark Side. But when she inspected her data more closely (as in the entry on *Orthogonality*), she saw that the pure-blooded wizards were often socially conservative – much more than magicians of mixed blood. In other words, BLOOD STATUS and CONSERVATISM overlapped (they showed an association), which made it harder to say which variable was *really* responsible for the allegiance that the wizards chose (Light or Dark). When a researcher believes (based on his or her analysis) that it was BLOOD STATUS, but it was actually CONSERVATISM, then the latter variable is a **confounder** (FIGURE 0.1).

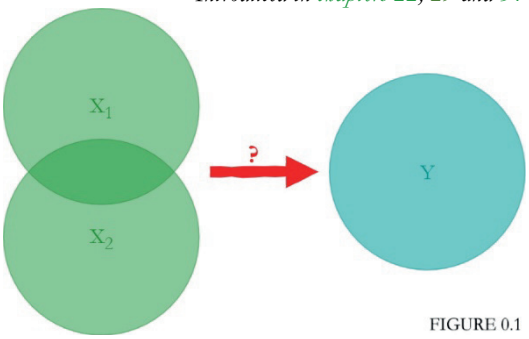


FIGURE 0.1

You should always try to correct for confounding and measure the 'pure' effect that each of the variables has. Multi-way ANOVA models and multiple regression models do this by default, as long as you include the potential confounder(s) in your analysis.

Three further notes:

- A confounder can also suppress the observed effect that of another independent variable. This means that said independent variable does have an effect, but the confounder makes that effect look weaker or even renders it invisible. Imagine that X_1 has a positive effect on the dependent variable Y, but X_2 has a negative effect; they may then cancel each other out.[3]
- Confounding requires overlap between the independent variables in order to occur. As such, it is only possible in non-orthogonal designs.
- Confounding is unrelated to interaction (see a later entry). Interaction can also exist as a real effect in the population, and it pops up in orthogonal designs as well as non-orthogonal ones. It implies that some x-variables *really* change each other's effects on the dependent variable Y. Notice how interaction always involves the dependent variable from the start, whereas the risk of confounding rises from overlap between the *independent* ones.

[3] This happens when X_1 is positively correlated with X_2 in addition.

CORRELATION

Introduced in chapter 3 (Parrrt 1)

When we wish to describe the linear relationship between two quantitative variables, we can make use of a correlation coefficient. This is a standardised measure; it always takes on a value between -1 and 1. FIGURE 0.2 helps you see what the value means. If r_{XY} is approximately 0, the relationship between the variables is very weak or even absent. The further r_{XY} deviates from 0, the stronger is the relationship. As I just stated it can be -1 at least; in that case we speak of a **perfect negative association**. If the correlation coefficient is 1 (maximum value), there's a **perfect positive association**.
You could calculate the linear correlation coefficient yourself, but you'll discover in the process why we let a computer do the work nowadays. 😐 The most common formula is:

$$r_{XY} = \frac{covariance_{XY}}{s_X s_Y} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})/(N-1)}{s_X s_Y}$$

This one applies to samples. I elaborate on it in chapter 3 in **Parrrt 1**.

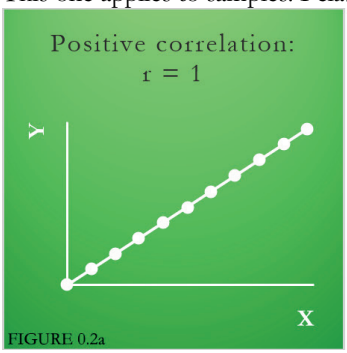


FIGURE 0.2a

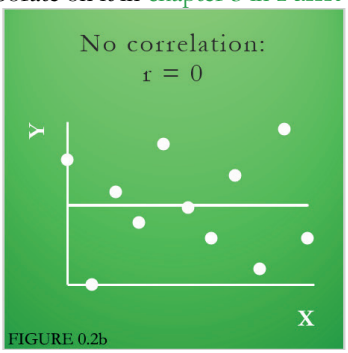


FIGURE 0.2b

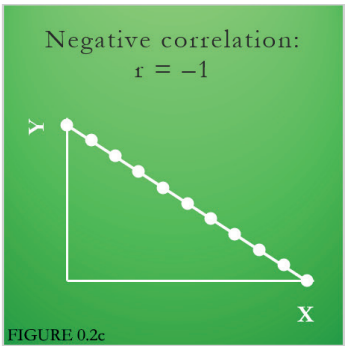


FIGURE 0.2c

DATA FORMAT

Not discussed elsewhere

Are you doing your own research? Make sure you set up the data file correctly. If you do this from the beginning, it can save a lot of pain in the bums. Data sets can be formatted in a few ways, and this determines the kind of analyses you can subject the data to in software like SPSS. Each chapter in *PPP* mentions the right format for the accompanying SPSS analysis.

UNIVARIATE (LONG) FORMAT

PERSON	THREAT	RESCUE (Y)
1	1	97
1	2	91
1	3	94
2	1	74
… (etc.)	…	…

An example of a univariate data file would look like above. The first row contains the rescue score of person 1 in threat condition 1, the second row contains the rescue score of the same person in threat condition 2, and so on. Every participant thus occupies multiple rows, and there is one column for a single dependent variable. All basic between-subjects designs work with this format, because every person is only measured once.

MULTIVARIATE (WIDE) FORMAT

Multivariate data files are used in particular for repeated measures (within-subjects) designs. They emphasise that each participant was measured multiple times, by making it look as if he or she scored on multiple dependent variables:

PERSON	RESCUE: disaster (Y_1)	RESCUE: terrorists (Y_2)	RESCUE: aliens (Y_3)
1	97	91	94
2	74	49	42
… (etc.)	…	…	…

Here every participant has one row in which three scores are found, one per 'dependent' variable (there are three, the number of columns). Note: the independent variable of the study, THREAT, has more or less been absorbed into the several 'dependent' variables *Y*.
Of course, multivariate data sets are also required for analyses that truly encompass multiple dependents, such as the MANOVA. Multi-level regression is the only major repeated measures technique that needs a univariate data format in SPSS.

WEIGHTED CASES FORMAT

FUTURE	count
1	90
2	18
3	12

Studies that only feature categorical variables can work with a shortened data set. You have to tell SPSS to weight each row using the 'count' variable (under <*Data*> <*Weight Cases*>). From then on, the software will see not just one person who scored future 1 (married), but 90 persons. In addition, there will be 18 persons with future 2 (divorced) and 12 with future 3 (dead).
Note that you can also process the same data in a univariate format, but then you'll need to make 90 rows with a 1 in it, 18 rows with a 2 in it, and so on.

DATA TRANSFORMATION

Introduced in chapter 1 (Parrrt 1)

Should someone measure, say, the length of some bananas in metres, he or she can choose to convert all scores to centimetres (which might also be a bit handier in this case). This conversion is known as a data transformation: all data are turned into other values, but their meaning stays the same. After all, the lengths of the bananas don't change if you go from 0,20 metres to 20 centimetres. The simplest kind of data transformations are **linear transformations**: multiplying, centring, and standardising (to z-scores; see further ahead). I feature them in section 1.6 of chapter 1 in **Parrrt 1**. Then there's also a more complicated set called **monotonic transformations**, such as taking the square root, taking the logarithm, and inversion (taking the reciprocal). Those are discussed in chapter 20.

DEGREES OF FREEDOM

Introduced in chapter 7 (Parrrt 1)

I'm very happy with my new explanation of degrees of freedom in chapter 7, section 7.2, if I may say so myself. It's easy and complete. ☺ I don't want to give you a watered-down summary, so just go there for a minute and read *Degrees of freedom explained*. Still seek a quicker impression? At your service:

- Degrees of freedom are the number of independent observations we can make when we calculate a certain sample estimate, or study the behaviour of a certain variable. Let's say you have a deck of 30 red and black cards, and you start counting the red ones. Your tally turns out to be 18 red cards; this is something you didn't know in advance. But now you *do* know at once that the number of black cards… must be 12. So as soon as you've tallied one colour, the number of cards from the other colour gets fixed. The colour variable, as we say, has only one degree of freedom.
- More degrees of freedom give the statistic or variable more 'wiggle room': it can change more when a new sample is drawn. Imagine introducing a third card colour.

- There are different kinds of degrees of freedom. Some of them depend on the size of the sample (N). These are always the degrees of freedom of the so-called 'error term' (ANOVA) or 'residual term' (regression). More degrees of freedom of this kind yield more power.

EFFECT SIZE

Introduced in chapter 6 (Parrrt 1)

When an effect is significant (which gives us evidence that it truly exists in the population), this doesn't tell the whole story. Is the effect pretty big or very small? This question is at least equally important. For instance, in chapter 7, the null hypothesis was rejected: it was found that women who use Pavlov Conditioner can indeed keep their hair in an attractive state for more than 48 hours… but how many? It turned out to be around 49 hours, which is hardly more than 48.
A myriad of quantities have been invented to express effect sizes. Many of them are unfortunately so abstract that they lose most of their real-world meaning. I'm a big fan of keeping things simple. Here are some easy indicators of effect size (they're not all flawless, but at least they're clear):

MEASURE	USE	ESTIMATE FORMULA	GUIDELINES
Cohen's *d* Standardised difference between means (z-score)	T-tests (to compare 2 groups at a time)	Depends on the research design. Basic setup: $\hat{d} = \frac{\bar{Y}_1 - \bar{Y}_2}{s}$	♦ 0,20: small ♦ 0,50: medium ♦ 0,80: large
Eta squared Proportion of explained variation	ANOVA	$\eta^2 = \frac{SS_{effect}}{SS_{(corrected)\ total}}$	♦ 0,01: small ♦ 0,06: medium ♦ 0,14: large
Odds ratio Ratio between two odds	Contingency tables, logistic regression	$OR = \frac{odds_{X=1}}{odds_{X=0}}$	♦ 1,68 (0,60): small ♦ 3,47 (0,29): medium ♦ 6,71 (0,15): large
Correlation Strength of linear relationship	Linear regression	Not worth your bother (it's complicated). Still want to know? See *Correlation* (earlier).	♦ 0,10: small ♦ 0,30: medium ♦ 0,50: large
R squared Proportion of explained variation	Linear regression	$R^2 = \frac{SS_{regression}}{SS_{total}}$	♦ 0,01: small ♦ 0,09: medium ♦ 0,25: large

---
EFFICIENT ESTIMATOR
>>> *Estimate*

ESTIMATE

Introduced in chapter 5 (Parrrt 1)

A quantity, such as a mean, a standard deviation, or a correlation, calculated for a sample. We use this sample estimate for estimating the corresponding population parameter. A few examples:

- A sample mean, $\bar{X}$, is used to estimate the population mean μ;
- A sample correlation, r_{XY}, is used to estimate the correlation ρ_{XY} at population level;
- Using a sample regression coefficient b, we estimate the population coefficient β.

As you can see, a Latin symbol always denotes an estimate in this book.
Since the value of an estimate can change whenever we draw a new sample, each estimate is a random variable with its own **sampling distribution**.

- An estimate is unbiased if its expected value equals the parameter. So, on average it should equal the parameter in case we drew the sample an infinite amount of times. Please note: you don't need an accurate (efficient) estimate for this!

INDEX

This list indicates the number of the section where the concept is introduced and/or covered extensively. You can always find it in the text by looking for a **bold term**. Concepts marked green are also explained in the *Overviews and Core Concepts* section. Some concepts labelled with *Overviews* are exclusively featured there.

1,5*IQR criterion	1.5, 20.1
-2 log likelihood	32-A.2, 33-A.4
Absolute frequencies	2.2
Addition rule	4.3
Additive model	31-B.1
Adjusted Bonferroni correction	21.2
Adjusted R Squared	15.4, 31-C.3
Akaike Information Criterion (AIC)	33-A.4
Allocation ratio	34.5
Alternative hypothesis	6.2
ANACOVA	*ANCOVA* (below)
Analysis of Covariance	Ch. 23, Ch. 27
Analysis of Variance	10.3
Anchor point	20.3
ANCOVA	Ch. 23, Ch. 27
Area under the ROC curve (AUC)	37.3
ARMA model	33-B.3 (end)
Association (categorical)	Ch. 2
Association (quantitative)	Ch. 3
Assumptions	7.3
Attenuation effect	36.7 (also 27.3)
Attenuation formula	36.7
Attributable risk	2.3
Autoregressive (marginal model)	33-B.3.V
Backward regression	31-A.1.III
Balanced design	22.4
Bar chart	1.3
Baron-Kenny method	31-C.1
Bartlett's test of sphericity	28.4.IV, 35.6
Base (logarithms)	29.6
Bayesian Information Criterion (BIC)	33-A.4
Between-subjects design	8.2
Bimodal distribution	1.4
Binary logistic regression	Ch. 32
Binomial distribution	4.5
Bivariate distribution	2.2
Bonferroni correction	10.6, 21.2.I
Box's test	26.3, 28.3
Box plot	1.5
Bozdogan's Criterion	33-A.4
Carry-over effects	24.3
Categorical variable	1.2
Central limit theorem	5.3
Central tendency	1.4
Centred leverage value	31-A.5
Centring	1.6
Chi-squared test	χ^2 (below)
Classification table	37.2
Coding scheme	30.3
Coefficient of determination	3.5, *R-squared* (below)
Cohen's kappa	16.2
Cohen's *d*	6.2
Collinearity	31-A.5
Column vector	28.3
Communality	35.4
Complement rule	4.3
Complete mediation	31-C.2
Compound Symmetry	33-B.3.III
Concurrent validity	37.4
Conditional distribution	2.2
Conditional main effect	31-B.3
Conditional probability	4.3
Confidence interval	6.4
Confirmatory factor analysis	35.3
Confounding	22.5, 29.5, 31-A.1
Constant factors	10.3
Construct validity	37.4
Content validity	37.4
Contingency table	2.2
Continuous distribution	5.3
Contrast coding (dummies)	30.3.III
Contrast analysis	21.3
Contrast coefficient	21.3
Contrast	21.3
Convergent validity	37.4
Cook's distance	31-A.5
Correlation matrix	35.2
Correlation	3.2, 15.4
Counterbalancing	24.3
Covariance matrix	26.3, 28.3, 35.2
Covariance structure	33-B.3
Covariance	3.2
Covariate	23.2, 23.3
Criterion validity	37.4
Critical area	6.2
Critical sample mean	6.5
Critical z-score	6.2
Cronbach's alpha	36.4
Cross tabulation	2.2
Cross-level interaction	33-B.6
Cubic contrast	21.3
Curvilinearity	3.2
Cut-off point (diagnostics)	37.2
Data equivalence	17.2, 25.4, 37.2
Data format	*Overviews*
Data reduction	35.2
Data transformation	1.6
Degrees of freedom	7.2
Density curve	5.4
Dependent t-test	Ch. 8
Deviance	33-A.4
Diagnostics	Ch. 37
Diagonal matrix	28.4.III
Diagonal model	33-B.3 (end)
Dichotomous variable	1.2
Difference score	8.3
Direct effect	31-C.2
Discrete distribution	5.3

Discriminant analysis 28.5
Discriminant function 28.5
Disjoint events 4.4
Dispersion 1.5
Distribution of sample scores 5.2
Divergent validity 37.4
Drawing with replacement 4.2
Drawing without replacement 4.2
Dummy variables Ch. 30
Durbin-Watson 31-A.4
e (mathermatical number) 29.6
Effect coding (dummies) 30.3.II
Effect modification *Interaction* (below)
Effect size 6.2
Efficient estimator 5.2
Eigenvalue 28.4.III, 35.4
Elementary outcome 4.2
E-matrix (Error matrix) 28.4.I
Epsilon correction 24.3, 24.5
Error 10.3
Error variance 10.5 (also 36.2)
Estimate 5.2
Estimated marginal mean 22.6, 23.5
Eta squared 10.5, 22.6
Etiologic fraction 2.3
Event 4.2
Expected agreement 16.2
Expected count 11.3, 12.3, 16.2
Expected value 4.5, 5.2
Experiment 23.3.II
Exploratory factor analysis 35.3
Face validity 37.4
Factor analysis Ch. 35
Factor loading 35.3
Factor pattern 35.3
Factor rotation 35.5
Factor 35.2
Factorial ANOVA Ch. 22
Factorial design 22.2
False negative 37.2
False positive 37.2
Family-wise error rate 10.6
F-change test 31-A.3
Fixed factor 24.2
Fixed part (regression model) 33-A.1
Forced-entry regression 31-A.1.II
Forward regression 31-A.1.III
F-ratio 10.5
Frequency table 1.3
Gain score analysis 27.3.II
Games-Howell correction 21.2
Glass' Δ (delta) 8.3, 9.4
G-matrix (Group matrix) 33-A.2
Goodness of fit 32-B.1.III, 33-A.4
Greenhouse-Geisser epsilon 24.5
Group centroid 28.5
Group factor 10.3
Growth curve model 33-B.5
HE^{-1} matrix 28.4.II
Hierarchical regression 31-A.1.I
Histogram 1.3
H-matrix (Hypothesis matrix) 28.4.I
Homoscedasticity 15.5, 31-A.4
Hosmer and Lemeshow test 32-A.3
Hotelling's trace 28.4.IV
Hurvich and Tsai's Criterion 33-A.4
Huynh-Feldt epsilon 24.5
Hypothesis test 6.2
Identifiability 33-A.7
Independent samples t-test Ch. 9
Indicator variable 30.3
Indirect effect 31-C.2
Influential case 31-A.5
Interaction effect 22.2, 29.4, 31-B.1
Intercept 3.3, 15.3
Inter-item correlation 36.2
Internal consistency 36.4
Interquartile range 1.5, 20.1
Inter-rater reliability Ch. 16
Interval variable 1.2
Intra-class correlation 33-A.2
Inverse transformation 20.3
Item p-value 36.5
Item 35.2, 36.2
Item-rest correlation 36.5
Item-total correlation 36.5
K1 rule 35.4
Kaiser criterion 35.4
Kaiser-Meyer-Olkin measure 35.6
Kappa 16.2
KMO 35.6
Kolmogorov-Smirnov test 20.2
Kruskal-Wallis test *Non-parametric*
Kurtosis 20.2
Least squares criterion 3.3
Least squares regression Ch. 15, Ch. 31
Level of measurement 1.2
Level-1 predictor 33-A.3
Level-2 predictor 33-A.3
Levene's test 9.3
Likelihood ratio test 32-A.2, 33-A.4
Likert scale 1.A
Linear contrast 21.3
Linear regression analysis 15.2
Linear transformation 1.6, *Data transformation*
Linear trend 3.2
Linear weighting (kappa) 16.2
Linearity assumption 23.4, 15.5, 31-A.4
Listwise deletion 25.4
Logarithm 29.6
Logarithmic transformation 20.3, 29.6
Logistic function 32-A.2
Logistic regression Ch. 32
Logit regression Ch. 32
Log-link (logit-link) 32-A.1
Long data format *Data format*
Lord's paradox 27.3.IV
Lower-bound (epsilon) 24.5
Main effect 22.2, 29.5, 31-B.1
Mann-Whitney U test *Non-parametric*
MANOVA Ch. 28
Mantel-Haenszel common OR 29.5
Marginal (distribution) 2.2
Marginal model 33-B.3
Matched pairs design 8.2
Matched-pairs t-test Ch. 8
Mauchly's test of sphericity 24.3
Maximum likelihood estimation 33-A.4
McNemar's test 14.3
Mean square 10.5
Mean 1.4
Measurement error 36.2
Median 1.4
Mediation Ch. 31-C

Mediator 31-C.1
Mixed ANOVA Ch. 26
Mixed model 24.2
Mode 1.4
Moderation *Interaction* (above)
Modern psychometrics 36.2
Monotonic transformation 20.3
Multicollinearity 31-A.5
Multi-level model 33-A.2
Multi-level regression Ch. 33
Multiple comparisons 10.6, 21.2
Multiple correlation 31-A.3
Multiple linear regression Ch. 31
Multiple observations per cell 24.4
Multiplicative model 31-B.1
Multivariate analysis of variance Ch. 28
Multivariate data format *Data format*
Multivariate normality 28.3
Multi-way ANOVA Ch. 22
Nested design 33-A.1
Nominal variable 1.2
Noncentrality parameter 34.5
Non-interaction assumption 23.4
Non-parametric test *Overviews*
Normal distribution 5.3
Normality test 20.2
Null hypothesis 6.2
Oblimin rotation 35.6
Oblique rotation 35.5
Observed agreement 16.2
Observed count 11.3, 12.3
Observed z-score 6.2
Odds ratio 29.3
Odds 29.3
One observation per cell 24.4
One sample t-test Ch. 7
One-sided/one-tailed test 6.2, 6.3
One-way ANOVA Ch. 10
Ordinal logistic regression 32-A.3
Ordinal variable 1.2
Ordinary least squares (OLS) Ch. 15, Ch. 31
Orthogonal contrasts 21.3
Orthogonal rotation 35.5
Orthogonality 22.4
Outlier in the x-direction 31-A.5
Outlier in the y-direction 31-A.5
Outlier 1.4, 20.1
Overfitting 3.6.I
Paired samples t-test Ch. 8
Pairwise comparisons 10.6, 21.2
Parallel items 36.2
Parallel-forms reliability 36.4
Parameter 5.2
Parsimony 32-B.1.III, 33-A.4
Part correlation 31-A.3
Part regression coefficient 31-A.3
Partial correlation 31-A.3
Partial eta squared 22.6
Partial mediation 31-C.2
P-hacking 30.3.III
Pie chart 1.3
Pillai's trace 28.4.IV
Planned comparison 21.3
Point of reference (ANCOVA) 23.5
Point-biserial correlation 30.2
Polynomial contrast 21.3
Pooled proportion 14.4
Pooled standard deviation 9.4
Pooled variance 10.4
Population distribution 5.2
Population 5.2
Power analysis Ch. 34
Power 6.5
Predictive capacity 15.4
Predictive validity 37.4
Principal axis factoring 35.3
Principal component analysis 35.3
Principal factor analysis 35.3
Probability distribution 4.2
Probability experiment 4.2
Probability 4.3
PROCESS (macro) 31-C.1
Product rule 4.3
Promax rotation 35.6
Proportion 13.2
Proportion of explained variation 3.4, 10.5, 15.4
Pseudo-R^2 32-A.2
Psychometrics Ch. 35-37
P-value 6.2
Q-Q (quantile-quantile) plot 20.2
Quadratic contrast 21.3
Quadratic weighting (kappa) 16.2
Quantitative variable 1.2
Quartile 1.5, 20.1
Quasi-experiment 23.2, 23.3.I
Random factor 24.2
Random intercept 33-A.2
Random part (regression model) 33-A.1
Random sample 4.2
Random slope 33-A.3
Random variable 4.2
Rank sum *Non-parametric*
Ratio variable 1.2
Reciprocal transformation 1.7
Reference coding (dummies) 30.3.I
Regression coefficients 3.3, 15.3
Regression equation 3.3
Regression line 3.3, 15.3
Relative frequencies 2.2
Relative risk 2.3,, 29.8
Reliability Ch. 36
Remaining/residual factors 10.3
Repeated measures ANOVA Ch. 24
Repeated measures design 8.2, 24.2
Reproduced correlation 35.4
Residual correlation 35.4
Residual plot 3.3
Residual 3.3, 15.3
Restricted Maximum Likelihood 33-A.4
Restriction of range 3.6
Reverse-coding 35.2
Risk difference 2.3
R-matrix (Residual matrix) 33-A.1
Robust 7.3
ROC curve 37.3
Root-n law 5.2
Roy's largest root 28.4.IV
R-squared 31-A.3
Sample space 4.2
Sample 5.2
Sampling distribution 5.2
Saturated model 33-B.2
Scaled Identity 33-B.3.I
Scatterplot 3.2, 15.2

Schwarz's Bayesian Criterion 33-A.4
Scree plot 35.4
Semi-partial correlation 31-A.3
Semi-partial regression coefficient 31-A.3
Sensitivity 37.2
Shapiro-Wilk test 20.2
Sign test *Non-parametric*
Significance level 6.2
Significance test 6.2
Simple effect 22.2
Simple regression Ch. 15
Simple slope 31-B.3
Simple structure 35.5
Simpson's paradox 2.B
Simultaneous confidence intervals 21.2
Simultaneous distribution 2.2
Skewness 20.2
Slope 3.3, 15.3
Spaghetti plot 33-A.3
Spearman-Brown formula 36.3
Specificity 37.2
Split-half reliability 36.4
Split-plot ANOVA Ch. 26
Split-plot design 26.2
Square root transformation 20.3
SSCP matrix 28.4.I
Standard deviation 1.5
Standard error of measurement 36.6
Standard error of the estimate 15.4
Standard error 5.2
Standardised regression coefficient 31-A.3
Standardising 1.6
Statistical independence 4.4
Stem-and-leaf plot 1.3
Stepwise regression 31-A.1.III
Stratified contingency table 29.4
Strong hypothesis 21.3
Structural equation modelling 31-C.1
Student's t-test 9.4
Studentised residual 31-A.5
Subgroups 1.4, 3.2
Success (mathematical) 13.3
Sum of cross products 3.2, 28.4
Sum of squares 1.5, 3.4, 10.2
Sum score 36.3
Suppressor variable 31-A.3
Test theory Ch. 35-37
Test-retest reliability 36.4
Three-level model 33-B.8
T-matrix (Total matrix) 28.4.I
Toeplitz model 33-B.3 (end)
Tolerance 31-A.5
Total effect 31-C.2
Treatment effect 10.3
Treatment variance 10.5
Tree diagram 4.5
Trimming 20.1
True negative 37.2
True positive 37.2
True score model 36.2
True score 36.2
Truncating 20.1
T-test 7.2
Tukey's HSD correction 21.2
Two samples t-test Ch. 9
Two-sided/two-tailed test 6.2, 6.3
Two-way ANOVA Ch. 22
Two-way within-subjects ANOVA Ch. 25
Type I error 6.5
Type II error 6.5
Type III sum of squares 22.5
Unbiased estimator 5.2
Unconditional probability 4.3
Uniform probability model 4.2
Unique variation 22.5
Uniqueness 35.4
Univariate data format *Data format*
Univariate distribution 1.3
Unpaired t-test Ch. 9
Unstructured model 33-B.3.II
Validity Ch. 37
Variable 1.2
Variance inflation factor (VIF) 31-A.5
Variance 1.5
Variate 28.4.III
Variation 1.5, 10.2
Varimax rotation 35.6
V-matrix (Variance matrix) 33-A.2
Wald test 32-A.2
Weighted mean (random variable) 4.5
Weighted variance (random variable) 4.5
Welch's ANOVA 10.5
Welch's t-test 9.4
Wide data format *Data format*
Wilcoxon rank-sum test *Non-parametric*
Wilcoxon signed-rank test *Non-parametric*
Wilks' lambda 28.4.IV
Winsorising 20.1
Within-subjects ANOVA Ch. 24
Within-subjects contrast 21.3
Within-subjects design 8.2
Youden index 37.3
Zero-order correlation 31-A.3
Z-score 1.6
Z-table 5.4
Z-test Ch. 6
Z-test for 1 proportion Ch. 13
Z-test for 2 proportions Ch. 14
χ^2-test for contingency tables Ch. 12
χ^2-test for equality of distributions Ch. 12
χ^2-test for goodness of fit Ch. 11
χ^2-test for independence Ch. 12