

SECOND EDITION



Essential Statistics for the Pharmaceutical Sciences

Philip Rowe



WILEY

Essential Statistics for the Pharmaceutical Sciences

Essential Statistics for the Pharmaceutical Sciences

Second Edition

Philip Rowe

Liverpool John Moores University, UK

WILEY

This edition first published 2016 © 2016 by John Wiley & Sons, Ltd.

Registered Office

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

Editorial Offices

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

111 River Street, Hoboken, NJ 07030-5774, USA

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

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

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

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Rowe, Philip, author.

Essential statistics for the pharmaceutical sciences / Philip Rowe. – Second edition.

p. ; cm.

Includes index.

ISBN 978-1-118-91338-3 (cloth) – ISBN 978-1-118-91339-0 (pbk.)

I. Title.

[DNLM: 1. Research Design. 2. Statistics as Topic. 3. Pharmacology--methods. QV 20.5]

RS57

615'.1072--dc23

2015015316

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

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

Cover images: © Ma-k/iStockphoto, © FotografiaBasica/iStockphoto, © Polygraphus/iStockphoto

Set in 10.5/12.5pt Minion by SPi Global, Pondicherry, India

To

Carol, Joshua and Nathan

for continued support

Contents

Preface	xiii
Statistical packages	xix
About the website	xxi

PART 1 PRESENTING DATA	1
-------------------------------	----------

1 Data types	3
1.1 Does it really matter?	3
1.2 Interval scale data	4
1.3 Ordinal scale data	4
1.4 Nominal scale data	5
1.5 Structure of this book	6
1.6 Chapter summary	6
2 Data presentation	7
2.1 Numerical tables	8
2.2 Bar charts and histograms	9
2.3 Pie charts	14
2.4 Scatter plots	16
2.5 Pictorial symbols	21
2.6 Chapter summary	22

PART 2 INTERVAL-SCALE DATA	23
-----------------------------------	-----------

3 Descriptive statistics for interval scale data	25
3.1 Summarising data sets	25
3.2 Indicators of central tendency: Mean, median and mode	26
3.3 Describing variability – Standard deviation and coefficient of variation	33
3.4 Quartiles – Another way to describe data	36
3.5 Describing ordinal data	40
3.6 Using computer packages to generate descriptive statistics	43
3.7 Chapter summary	45
4 The normal distribution	47
4.1 What is a normal distribution?	47
4.2 Identifying data that are not normally distributed	48
4.3 Proportions of individuals within 1SD or 2SD of the mean	52

4.4	Skewness and kurtosis	54
4.5	Chapter summary	57
4.6	Appendix: Power, sample size and the problem of attempting to test for a normal distribution	58
5	Sampling from populations: The standard error of the mean	63
5.1	Samples and populations	63
5.2	From sample to population	65
5.3	Types of sampling error	65
5.4	What factors control the extent of random sampling error when estimating a population mean?	68
5.5	Estimating likely sampling error – The SEM	70
5.6	Offsetting sample size against SD	74
5.7	Chapter summary	75
6	95% Confidence interval for the mean and data transformation	77
6.1	What is a confidence interval?	78
6.2	How wide should the interval be?	78
6.3	What do we mean by '95%' confidence?	79
6.4	Calculating the interval width	80
6.5	A long series of samples and 95% C.I.s	81
6.6	How sensitive is the width of the C.I. to changes in the SD, the sample size or the required level of confidence?	82
6.7	Two statements	85
6.8	One-sided 95% C.I.s	85
6.9	The 95% C.I. for the difference between two treatments	88
6.10	The need for data to follow a normal distribution and data transformation	90
6.11	Chapter summary	94
7	The two-sample t-test (1): Introducing hypothesis tests	95
7.1	The two-sample t -test – an example of an hypothesis test	96
7.2	Significance	103
7.3	The risk of a false positive finding	104
7.4	What aspects of the data will influence whether or not we obtain a significant outcome?	106
7.5	Requirements for applying a two-sample t -test	108
7.6	Performing and reporting the test	109
7.7	Chapter summary	110
8	The two-sample t-test (2): The dreaded P value	111
8.1	Measuring how significant a result is	111
8.2	P values	112
8.3	Two ways to define significance?	113
8.4	Obtaining the P value	113
8.5	P values or 95% confidence intervals?	114
8.6	Chapter summary	115

9	The two-sample <i>t</i>-test (3): False negatives, power and necessary sample sizes	117
9.1	What else could possibly go wrong?	118
9.2	Power	119
9.3	Calculating necessary sample size	122
9.4	Chapter summary	130
10	The two-sample <i>t</i>-test (4): Statistical significance, practical significance and equivalence	131
10.1	Practical significance – Is the difference big enough to matter?	131
10.2	Equivalence testing	135
10.3	Non-inferiority testing	139
10.4	<i>P</i> values are less informative and can be positively misleading	141
10.5	Setting equivalence limits prior to experimentation	143
10.6	Chapter summary	144
11	The two-sample <i>t</i>-test (5): One-sided testing	145
11.1	Looking for a change in a specified direction	146
11.2	Protection against false positives	148
11.3	Temptation!	149
11.4	Using a computer package to carry out a one-sided test	153
11.5	Chapter summary	153
12	What does a statistically significant result really tell us?	155
12.1	Interpreting statistical significance	155
12.2	Starting from extreme scepticism	159
12.3	Bayesian statistics	160
12.4	Chapter summary	161
13	The paired <i>t</i>-test: Comparing two related sets of measurements	163
13.1	Paired data	163
13.2	We could analyse the data by a two-sample <i>t</i> -test	165
13.3	Using a paired <i>t</i> -test instead	165
13.4	Performing a paired <i>t</i> -test	166
13.5	What determines whether a paired <i>t</i> -test will be significant?	169
13.6	Greater power of the paired <i>t</i> -test	170
13.7	Applicability of the test	170
13.8	Choice of experimental design	171
13.9	Requirement for applying a paired <i>t</i> -test	172
13.10	Sample sizes, practical significance and one-sided tests	173
13.11	Summarising the differences between paired and two-sample <i>t</i> -tests	175
13.12	Chapter summary	175
14	Analyses of variance: Going beyond <i>t</i>-tests	177
14.1	Extending the complexity of experimental designs	177
14.2	One-way analysis of variance	178
14.3	Two-way analysis of variance	188

14.4	Fixed and random factors	198
14.5	Multi-factorial experiments	204
14.6	Chapter summary	204
15	Correlation and regression – Relationships between measured values	207
15.1	Correlation analysis	208
15.2	Regression analysis	218
15.3	Multiple regression	225
15.4	Chapter summary	235
16	Analysis of covariance	237
16.1	A clinical trial where ANCOVA would be appropriate	238
16.2	General interpretation of ANCOVA results	239
16.3	Analysis of the COPD trial results	241
16.4	Advantages of ANCOVA over a simple two-sample <i>t</i> -test	244
16.5	Chapter summary	249
PART 3	NOMINAL-SCALE DATA	251
17	Describing categorised data and the goodness of fit chi-square test	253
17.1	Descriptive statistics	254
17.2	Testing whether the population proportion might credibly be some pre-determined figure	258
17.3	Chapter summary	264
18	Contingency chi-square, Fisher's and McNemar's tests	265
18.1	Using the contingency chi-square test to compare observed proportions	266
18.2	Extent of change in proportion with an expulsion – Clinically significant?	270
18.3	Larger tables – Attendance at diabetic clinics	270
18.4	Planning experimental size	273
18.5	Fisher's exact test	275
18.6	McNemar's test	277
18.7	Chapter summary	279
18.8	Appendix	280
19	Relative risk, odds ratio and number needed to treat	283
19.1	Measures of treatment effect – relative risk, odds ratio and number needed to treat	283
19.2	Similarity between relative risk and odds ratio	287
19.3	Interpreting the various measures	288
19.4	95% confidence intervals for measures of effect size	289
19.5	Chapter summary	293
20	Logistic regression	295
20.1	Modelling a binary outcome	295
20.2	Additional predictors and the problem of confounding	304

20.3	Analysis by computer package	307
20.4	Extending logistic regression beyond dichotomous outcomes	308
20.5	Chapter summary	309
20.6	Appendix	309
PART 4	ORDINAL-SCALE DATA	311
21	Ordinal and non-normally distributed data: Transformations and non-parametric tests	313
21.1	Transforming data to a normal distribution	314
21.2	The Mann–Whitney test – a non-parametric method	318
21.3	Dealing with ordinal data	323
21.4	Other non-parametric methods	325
21.5	Chapter summary	333
21.6	Appendix	334
PART 5	OTHER TOPICS	337
22	Measures of agreement	339
22.1	Answers to several questions	340
22.2	Several answers to one question – do they agree?	344
22.3	Chapter summary	358
23	Survival analysis	361
23.1	What special problems arise with survival data?	362
23.2	Kaplan–Meier survival estimation	363
23.3	Declining sample sizes in survival studies	369
23.4	Precision of sampling estimates of survival	369
23.5	Indicators of survival	371
23.6	Testing for differences in survival	374
23.7	Chapter summary	383
24	Multiple testing	385
24.1	What is it and why is it a problem?	385
24.2	Where does multiple testing arise?	386
24.3	Methods to avoid false positives	388
24.4	The role of scientific journals	392
24.5	Chapter summary	393
25	Questionnaires	395
25.1	Types of questions	396
25.2	Sample sizes and low return rates	398
25.3	Analysing the results	399
25.4	Problem number two: Confounded questionnaire data	401
25.5	Problem number three: Multiple testing with questionnaire data	401
25.6	Chapter summary	403
Index		405

Preface

At whom is this book aimed?

Statisticians or statistics users?

The starting point for writing this book was my view that most existing statistics books place far too much emphasis on the mechanical number crunching of statistical procedures. This makes the subject seem extremely tedious and (more importantly) diverts attention from what are actually vital and interesting fundamental concepts. I believe that we need to distinguish between ‘Statisticians’ and ‘Statistics users’. The latter are the people at whom this book is aimed – those thousands of people who have to use statistical procedures without having any ambition to become statisticians.

There is any number of student programmes which include an element of statistics. These students will have to learn to use at least the more basic statistical methods. There are also those of us engaged in research in academia or industry. Some of us will have to carry out our own statistical analyses and others will be able to call on the services of professional statisticians. However, even where professionals are to hand, there is still the problem of communication; if you don’t even know what the words mean, you are going to have great difficulty explaining to a statistician exactly what you want to do. The intention is that all of the above should find this book useful.

As a statistics user, what you really need to know is:

- Why are statistical procedures necessary at all?
- How can statistics help in planning experiments?
- Which procedure should I employ to analyse the results?
- What do the statistical results actually mean when I’ve got them?

This book is quite happy to treat any statistical calculation as a black box. It will explain what needs to go into the box and it will explain what comes out the other end. But do you really need to know what goes on inside the box? This approach isn’t just lazy or negative. By stripping away all the irrelevant bits, we can focus on the aspects that actually matter. This book will try to concentrate on the issues listed above – the things that statistics users really do need to understand.

To what subject area is the book relevant?

All the procedures and tests are illustrated with practical examples and data sets. The cases are drawn from the pharmaceutical sciences and this is reflected in the book's title. However, pretty well all the methods described and the principles explored are perfectly relevant to a wide range of scientific research, including pharmaceutical, biological, biomedical and chemical sciences.

At what level is it aimed?

The book is aimed at everybody from undergraduate science students and their teachers to experienced researchers.

The first few chapters are fairly basic. They cover data description (mean, median, mode, standard deviation and quartile values) and introduce the problem of describing uncertainty due to sampling error (Standard Error of the Mean and 95% Confidence Interval for the mean). These chapters are mainly relevant to first year students.

Later chapters then cover the most commonly used statistical tests with a general trend towards increasing complexity. The approach used is not the traditional one of giving equal weight to a wide range of techniques. As the focus of the book is the issues surrounding statistical testing rather than methods of calculation, one test (the two-sample t -test) has been used to illustrate all the relevant issues (Chapters 7–11). Further chapters then deal with other tests more briefly, referring back to general principles that have already been established.

What has changed since the first edition of this book in 2007?

My motivation for producing a second edition has very little to do with the arrival of any new statistical methods that are likely to have broad applicability for working pharmaceutical scientists – there are precious few.

So, why a new edition? I provide statistical advice to researchers in diverse areas of pharmaceutical science (and beyond) and the change I have noticed is an increased familiarity and confidence with the use of statistical packages. This brings both opportunities and pitfalls.

Opportunities

There are several statistical methods that I considered covering in the first edition but I concluded, at that time, that very few researchers would have the confidence to tackle them. Hopefully we have now moved on. For this edition I have added analysis of covariance, logistic regression, measures of agreement (e.g. Cronbach's Alpha

and Cohen's Kappa) and survival analysis. Many of these are more advanced than the topics in the first edition, but with some clear explanatory material (which I hope I have supplied) and relatively easy to use statistical packages, most pharmaceutical scientists should be perfectly capable of applying them.

Pitfalls

On the negative side, powerful statistical packages also offer new and improved methods to make a complete fool of yourself. Where I have seen examples of this over the last seven years I have tried to include warnings in this new edition.

Other new material

Apart from the completely new topics listed earlier, I have also filled in a number of gaps from the first edition. Many of these additions concern studies that generate simple dichotomous outcomes (e.g. Yes/No or Success/Failure). I have added the use of the Relative Risk, Odds Ratio and Number Needed to Treat (RR, OR and NNT) as descriptors of the extent of change in a dichotomous outcome. I have also described Fisher's and McNemar's tests as additions to the simple chi-square test which was included in the first edition.

Finally, when you teach statistics to various groups of students, year in, year out, you inevitably have the occasional light-bulb moment, when you realise that there is actually a much better way to explain something than the awkward method you have used for the last 30 years. Some of these are scattered around the book.

Key point and pirate boxes

Key point boxes

Throughout the book you will find key point boxes that look like this:



Proportions of individuals within given ranges

For data that follows a normal distribution:

- About two-thirds of individuals will have values within 1 SD of the mean.
- About 95% of individuals will have values within 2 SD of the mean.

These never provide new information. Their purpose is to summarise and emphasise key points.

Pirate boxes

You will also find pirate boxes that look like this:



Switch to a one-sided test after seeing the results

Even today, this is probably the best and most commonly used statistical fiddle.

You did the experiment and analysed the results by your usual two-sided test. The result fell just short of significance (P somewhere between 0.05 and 0.1). There's a simple solution – guaranteed to work every time. Re-run the analysis, but change to a one-sided test, testing for a change in whatever direction you now know the results actually suggest.

Until the main scientific journals get their act into gear, and start insisting that authors register their intentions in advance, there is no way to detect this excellent fiddle. You just need some plausible reason why you 'always intended' to do a one-tailed test in this particular direction, and you're guaranteed to get away with it.

These are written in the style of Machiavelli, but are not actually intended to encourage statistical abuse. The point is to make you alert for misuses that others may try to foist upon you. Forewarned is forearmed.

The danger posed, is reflected by the number of skull and cross-bone symbols.



Minor hazard. Abuse easy to spot or has limited potential to mislead.



Moderate hazard. The well-informed (e.g. readers of this book) should spot the attempted deception.



Severe hazard. An effective ruse that even the best informed may suspect, but never be able to prove.

A potted summary of this book

The book is aimed at those who have to use statistics, but have no ambition to become statisticians *per se*. It avoids getting bogged down in calculation methods and focuses instead on crucial issues that surround data generation and analysis (Sample size estimation, interpretation of statistical results, the hazards of multiple

testing, potential abuses etc.). In this day of statistical packages, it is the latter that cause the real problems, not the number-crunching.

The book's illustrative examples are all taken from the pharmaceutical sciences, so students (and staff) in the areas of pharmacy, pharmacology and pharmaceutical science should feel at home with all the material. However, the issues considered are of concern in most scientific disciplines and should be perfectly clear to anybody from a similar discipline, even if the examples are not immediately familiar.

Material is arranged in a developmental manner. The first six chapters are fairly basic, with special emphasis on random sampling error. The next block of five chapters uses the two-sample t -test to introduce a series of general statistical principles. Remaining chapters then cover other topics in (approximately) increasing order of complexity.

The book is not tied to any specific statistical package. Instructions should allow readers to enter data into any package and find the key parts of the output. Specific instructions for performing all the procedures, using Minitab or SPSS, are provided in a linked website (www.ljmu.ac.uk/pbs/rowestats/).

Statistical packages

There are any number of statistical packages available. It is not the intention of this book to recommend any particular one.

Microsoft Excel

Probably the commonest way to collect data and perform simple manipulations is within a Microsoft Excel (XL) spreadsheet. Consequently, the most obvious way to carry out statistical analyses of such data would seem to lie within XL itself. Let me give you my first piece of advice. Don't even consider it! The data analysis procedures within XL are rubbish – a very poor selection of procedures, badly implemented. (Apart from that, they are OK.) It is only at the most basic level that XL is of any real use (calculation of the mean, SD and SEM). It is therefore mentioned in some of the early chapters but not thereafter.

Other packages

A decision was taken not to include blow by blow accounts of how to perform specific tests using any package, as this would excessively limit the book's audience. Instead, general comments are made about:

- Entering data into packages;
- The information that will be required before any package can carry out the procedure;
- What to look for in the output that will be generated.

The last point is usually illustrated by generic output. This will not be in the same format as that from any specific package, but will present information that they should all provide.

Detailed instructions for Minitab and SPSS on the website

As Minitab and SPSS clearly do have a significant user base, detailed instructions on how to use these packages to execute the procedures in this book are available through the website (www.ljmu.ac.uk/pbs/rowestats/). These cover how to:

- Arrange the data for analysis.
- Trigger the appropriate test.
- Select appropriate options where relevant.
- Find the essential parts of the output.

About the website

Supplementary material, including full data sets and detailed instructions for carrying out analyses using packages such as SPSS or Minitab, is provided at:

www.ljmu.ac.uk/pbs/rowestats/

Part 1

Presenting data

1

Data types

This chapter will ...

- Set out a system for describing different types of data.
- Explain why we need to identify the type of data with which we are dealing.

1.1 Does it really matter?

To open a statistics book with a discussion of the way in which data can be categorised into different types probably sounds horribly academic. However, the first step in selecting a data handling technique is generally identifying the type of data with which we are dealing. So, it may be dry, but it does have real consequences.

We will discuss three types of data. These go under a variety of names. The names that this book will use are (with common alternatives in brackets):

- Interval scale (Continuous measurement data)
- Ordinal scale (Ordered categorical data)
- Nominal scale (Categorical data)

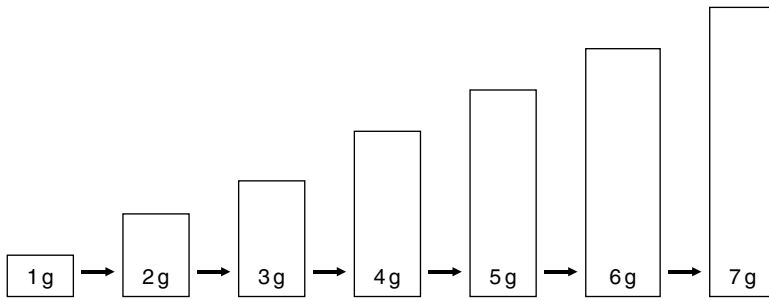


Figure 1.1 Interval scale data – a series of weights (1–7 g)

1.2 Interval scale data

The first two types of data that we will consider are both concerned with the measurement of some characteristic. ‘Interval scale’ (or what is sometimes called ‘Continuous measured’) data includes most of the information that would be generated in a laboratory. These include weights, lengths, timings, concentrations, pressures etc. Imagine we had a series of objects weighing 1, 2, 3 and so on up to 7 g, as in Figure 1.1.

Now think about the differences in weights as we step from one object to the next. These steps, each of one unit along the scale, have the following characteristics:

1. *The steps are of an exactly defined size.* If you told somebody that you had a series of objects like those described above, he or she would know exactly how large the weight differences are as we progressed along the series.
2. *All the steps are of exactly the same size.* The weight difference between the 1 and 2 g objects is the same as the step from 2 to 3 g or 6 to 7 and so on.

Because these measurements have constant sized steps (intervals), the measurement scale is described as a ‘Constant interval scale’ and the data as ‘Interval scale’. Although the weights quoted in Figure 1.1 are exact integers, weights of 1.5 or 3.175 g are perfectly possible, so the measurement scale is said to be ‘Continuous’.

1.3 Ordinal scale data

Again measurement is involved, but the characteristic being assessed is often more subjective in nature. It’s all well and good to measure nice neat objective things like blood pressure or temperature, but it’s also a good idea to get the patient’s angle on

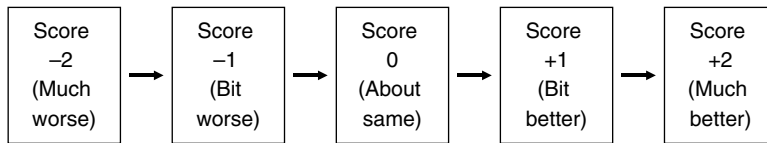


Figure 1.2 Ordinal scale data – scores for patient responses to treatment

how they feel about their treatment. The most obvious way to do this is as a score, of (say) -2 to $+2$ with the following equivalences:

- -2 = Markedly worse
- -1 = A bit worse
- 0 = About the same
- $+1$ = A bit better
- $+2$ = Markedly better

In this case (Figure 1.2) all we know is that if one patient reports a higher value than another, they are more satisfied with their outcome. However, we have no idea how much more satisfied he/she might be.

Since we have no idea how large the steps are between scores, we obviously could not claim that all steps are of equal size. In fact, it is not even necessarily the case that the difference between scores of -2 and 0 is greater than that between $+1$ and $+2$. So, neither of the special characteristics of a constant interval scale apply to this data.

The name ‘Ordinal’ reflects the fact that the various outcomes form an ordered sequence going from one extreme to its opposite. Such data is sometimes referred to as ‘Ordered categorical’. In this case the data is usually discontinuous; individual cases being scored as -1 or $+2$ and so on, with no fractional values.

1.4 Nominal scale data

In this case there is no sense of measuring a characteristic; we use a system of classifications, with no natural ordering. For example, one of the factors that influences the effectiveness of treatment could be the specific manufacturer of a medical device. So, all patients would be classified as users of ‘Smith’, ‘Jones’, or ‘Williams’ equipment. There is no natural sequence to these; they are just three different makes.

With ordinal data we did at least know that a case scored as (say) $+2$ is going to be more similar to one scored $+1$ than to one scored 0 or -1 . But, with nominal data, we have no reason to expect Smith or Jones equipment to have any special degree of similarity. Indeed the sequence in which one would list them may be entirely arbitrary.

Quite commonly there are just two categories in use. Obvious cases are Male/Female, Alive/Dead or Success/Failure. In these cases, the data is described as “Dichotomous”.



Data types

Interval scale: Measurements with defined and constant intervals between successive values. Values are continuous.

Ordinal scale: Measurements using classifications with a natural sequence (lowest to highest) but with undefined intervals. Values are discontinuous.

Nominal scale: Classifications that form no natural sequence.

1.5 Structure of this book

The structure of this book is largely based upon the different data types. Chapters 3 to 16 all deal with the handling of continuous measurement data, with Chapters 17 to 20 focusing on categorical data; and then Chapter 21 covers ordinal data.

1.6 Chapter summary

When selecting statistical procedures, a vital first step is to identify the type of data that is being considered.

Data may be:

- Interval scale: Measurements on a scale with defined and constant intervals. Data is continuous.
- Ordinal scale: Measurements on a scale without defined intervals. Data is discontinuous.
- Nominal scale: Classifications that form no natural sequence.