Applying Statistics in Behavioural Research

Applying Statistics in Behavioural Research

Jules L. Ellis

Inge M. Rabeling

Boom

visit the
website!

By registering the unique activation code below on www.boomstudent.nl you can gain access to the online book. This code is strictly personal and connected to the 2nd edition. After activation you will have access to the online material for the duration of two years. The code can be activated until six months before the arrival of a new edition. The code can only be used once. Are you a teacher? Then you have received your activation code via e-mail. This code can be activated on www.boomdocent.nl.

Cover design: Emma Raben

ISBN 9789024473298
NUR 916

www.boomstudent.nl
www.boom.nl/hogeronderwijs

Contents

Preface for students

Relevance of the subject

Psychology is based on human behaviour. It is therefore important to observe, with our own eyes, how people behave. The problem you encounter is that people display a great diversity and variety in behaviour—more than one person can possibly grasp. That is why researchers in psychology use statistics. Statistics helps with:

- systematically collecting and coding observations
- organising and summarising data
- interpreting the results
- communicating all of this to other researchers

In short: without statistics, there is no psychology.

Many psychology students dread the statistics courses in their programme. If you do not recognise this at all and are eager to get started with the content, you can skip the rest of this preface, and we wish you much pleasure and success with the course. If you *do* recognise this, we wish you the same, but we also hope that the following explanation will clear up a few persistent misconceptions and put you at ease.

Misconceptions about the content of statistics

Statistics aims to answer questions using data. Many people are insufficiently aware of this, which leads to several misconceptions. Before addressing the nature of these misconceptions, let us first consider their possible cause. One cause could be that many people think statistics is a form of mathematics. It is not. Statistics is about data. Mathematics is, in fact, the only science that does not use data. Mathematicians will never collect data in their lifetime. Admittedly, statistics uses a lot of mathematics, but it is not mathematics itself.

As a result of this misunderstanding, people sometimes complain that certain exercises are unclear. For example, an exercise might ask for a numerical summary, without defining the term in the book. That would be a valid criticism if statistics were a form of mathematics. But it is not. Statistics aims to answer questions using data, and such questions are often unclear. This is the reality in almost all research. This uncertainty is simply part of the process, and you must learn to deal with it. In the above example, you must decide for yourself what a numerical summary is. That is not so hard: it should contain numbers, and it should be a summary. It must be a *good* summary—one that answers the main questions people might ask. So *you* must determine which questions are important. In mathematics, you may never have had to do this, but in statistics it is essential.

Another misconception is that statistics is about numbers and calculations. You may see someone doing a lot of calculations and arriving at the result '42'—and then they stop. But statistics aims to answer questions using data. If the question is, for example, 'What do these data tell us about the differences between men and women?', then '42' is obviously not an answer. You must draw conclusions and formulate and write them down precisely. So a good answer is not '42' or '42 → difference' but rather: 'There is a difference of 42 points between Dutch men and women of middle age in their average stress tolerance, measured with the SQV-5.' You might think such a sentence is unimportant in an exam, but you could be in for a nasty surprise. Many statistics teachers consider this the most important part. In our exams, most people fail because they omit or incorrectly phrase the interpretations. After all, what is the point of calculating everything perfectly if you then draw the wrong conclusion—or no conclusion at all?

When doing exercises, keep in mind that communication and common sense are even more important in statistics than in everyday life. If this book does not explicitly state that you should report the unit of measurement, you might still consider whether doing so is a good way to communicate.

Myths about learning statistics

Over several years, we have analysed the exam results of students in our statistics courses. The conclusions were always the same. The main reason students failed was simply that they did not sit the exam at all. Exam grades in statistics were barely correlated with mathematics grades from secondary school, but strongly correlated with the amount of homework done. Roughly speaking, the probability of passing was about equal to the percentage of homework completed. Those who always did all the homework had a 100% chance of passing, either on the first attempt or the resit. Those who did 50% had about a 50% chance. Our message is simple: do all your homework and take the exam.

The message is simple and the argument clear, but some students still do not believe it. This may be because the argument itself is statistical. That is why we will now explain something about the psychology of learning statistics—which will, of course, require many more words.

One of the main reasons people fail statistics exams may be hard for you to believe: they have the wrong psychological theory. It is psychological, not statistical, because it concerns human thinking performance. The theory is: statistics is a matter of innate ability—you either have it or you do not. Before we address this theory, ask yourself to what extent you believe it, and why.

Research shows that those who believe this theory tend to perform poorly in statistics. By contrast, those who do well usually believe that their performance is mainly a matter of hard work and patience. In this respect, it is somewhat like mathematics. Mathematicians have an expression that you need 'sitzfleisch' (the ability to sit and persevere). Without it, you will not get far.

You do not need to be a great psychologist to see that the innate-ability theory is disastrous for someone's performance in statistics. If it is true, there is no point in studying—and so people do not. Then they fail. And then they say: 'See? I can't do it.' We have heard this before. If you believe in the innate-ability theory, you may still not be convinced otherwise. But why do you believe it? Perhaps in your class there were pupils for whom it all seemed effortless, and others who always had to slog away. Perhaps teachers or parents told you: 'That's not for you.' For many people, it is psychologically advantageous to maintain this myth. Those who supposedly find it effortless are proud of it, and therefore downplay the time they actually put in. Surprisingly, even those who had to slog are often proud of that, declaring while working on a problem: 'I'm a humanities person—I can't do this,' with a certain satisfaction.

Before you read on, take a minute to imagine how such a person approaches a statistics problem. A typical problem requires a few minutes of thought before you can start answering. Now picture someone secretly proud of not being able to do statistics. How will they approach a statistics problem?

A person like that will not succeed. After five seconds, they think: 'I don't know the answer. See? I'm a humanities person!'—which is exactly what they want to think about themselves. This thought would not be that bad if they would then just start solving the problem. However, the trouble is, they then stop working on the problem and expect someone else to help them, or they keep repeating such thoughts for minutes on end. The problem never gets solved, and everyone sees this as confirmation of the theory. But in reality, it is just a self-fulfilling prophecy.

There is another myth about being a 'humanities person': that a humanities person is not a science person. Dozens of studies have shown that all cognitive performances correlate positively: people who are good at humanities subjects are usually also good at science subjects, and vice versa. Yet the myth claims the opposite. Here too, a self-fulfilling prophecy arises: once someone concludes they are a humanities person and believes this is incompatible with science ability, they tend to neglect science subjects—thinking there is no point in trying.

Closely related to the innate-ability myth is this one: statistics is a matter of 'insight.' People who say this often secretly add: '...and you either have it or you don't,' which is just the innate-ability myth again. But even without that addition, it is a myth. Statistics, at least as taught in this book, is 70% practice, 30% memorisation, and 10% insight. Yes, you read that right—and you can do the maths. Insight is not something you start with, or a prerequisite; it is something you *end up with* as a result of practice and memorisation. What is 'insight' anyway? We have never heard a clear definition. Insight seems to be little more than well-organised knowledge and skills, plus a good feeling about them. This brings us to another related myth: memorisation is for the less intelligent. Quite a few students—especially those who want to be seen as smart—refuse to memorise things because they feel it is beneath them. Please believe us: you can only truly understand things once you know them largely by heart.

These myths share the assumption that performance in statistics is tied to unchangeable personal traits. They are often secret choices people make about what they want to believe about themselves. In doing so, they ignore the fact that statistics can be learned—just as one can learn to play the piano. But you must practise. No one, not even prodigies like Mozart, can play the piano at birth. It takes practice first, with 'blood, sweat, and tears'. The same is true for statistics.

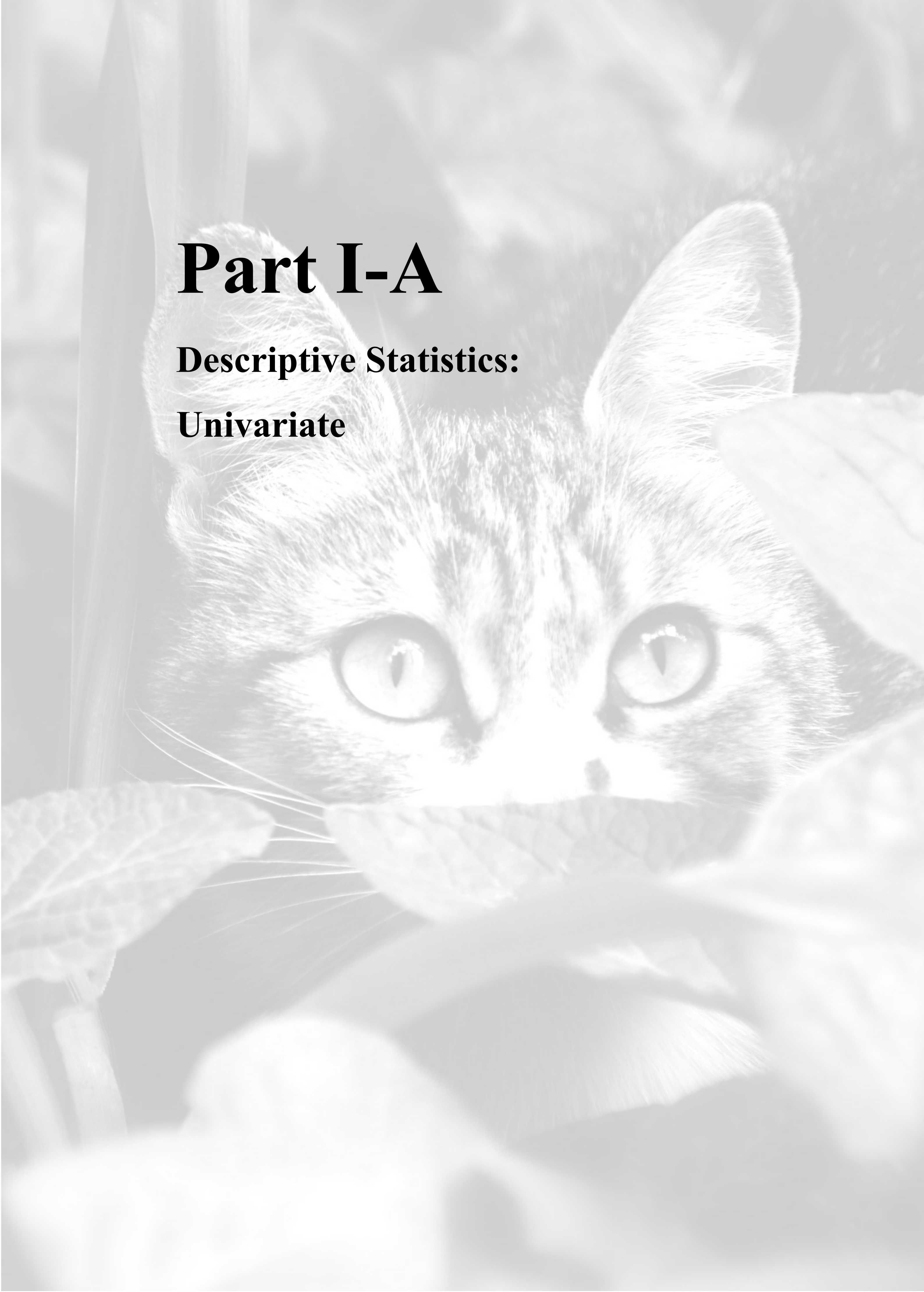
If, while studying this subject, you often find yourself thinking: 'I can't do this,' we encourage you to reframe that thought into: 'I can't do this *yet*.' Adopt a growth mindset: you can learn new things. And sometimes that means stepping outside your comfort zone. You didn't learn to read in a single day either—and in the end, you managed that too. But if, against our advice, you keep telling yourself that you will never succeed, then we have a challenge for you. This course is not going away. You cannot graduate without it. If the thought 'I can't pass this course' were really true, the consequence would be that you would never graduate. In that case, you might as well stop your studies right now, don't you think? Please realize that we do not expect everyone will or can become a Rachmaninoff of statistics. But a student who disciplines themselves to study regularly (without cramming only right before the exam), asks questions when things are unclear, and spends a lot of time working through problems, will find that a statistics course is perfectly manageable.

Jules L. Ellis & Inge M. Rabeling,
Nijmegen

Part I-A

Descriptive Statistics:

Univariate



1 Introduction of the examples

What is statistics? According to the American Statistical Association (2016),

> 'Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances (Davidian, M. and Louis, T. A., 10.1126/science.1218685).'

So, for learning about statistics then, we need data. Several data examples will be considered repeatedly throughout this book. These examples come from different behavioural research areas. In this chapter, the examples are introduced.

1.1 Example 1: Clinical psychology (depression prevention)

Reference of the study. This example is based on Veltman, Ruiter, and Hosman (1996).

Goal of the study. The study evaluated a training program aimed to prevent depression.

Subjects. Participants were individuals aged between 15 and 19 from the Dutch cities of Nijmegen and Arnhem who were believed to be at risk of developing a depression.

Intervention. Some participants were given a training to reduce automatic negative thoughts (ANTs). An example of an ANT is the thought 'what's wrong with me' if one cannot stop the thought from recurring again and again. Automatic negative thoughts are an important cause of depression according to 'Beck's cognitive theory of depression' (Moilanen, 1993). The training therefore was hoped to reduce the risk of depression. The goal of the study was to evaluate whether this was actually achieved. The participants who did not receive the training served as a **control** group. The training will henceforth be called the **treatment**.

Measures. The degree of depression was assessed by the Beck Depression Inventory (BDI), which contains 21 multiple choice questions. In each question, the participant has to pick one out of four possible alternatives. An example of one question is this:

- ☐ I do not feel sad
- ☐ I feel sad
- ☐ I am sad all the time and I can't snap out of it
- ☐ I am so sad or unhappy that I can't stand it

The answer of the participant is scored as 0, 1, 2, or 3. For example, if the participant has picked the third alternative ('I am sad all the time and I can't snap out of it'), then the participant's score for that question would be 2. After completing the questionnaire, the scores of the participant are summed, yielding the participant's total score.
A similar questionnaire was used to measure the amount of automatic negative thoughts of the participant. The frequency of such thoughts was assessed by the Automatic Thoughts Questionnaire (ATQ) (Hollon & Kendall, 1980), which has 30 questions. An example of one of those questions is:

	not at all	sometimes	moderately often	often	all the time
My life is a mess	1	2	3	4	5

Each question is scored by the indicated number. The scores are summed per participant.

Repeated measures. Both the BDI and the ATQ were administered three times to each participant: before the treatment period, after the treatment period and after a follow-up. Only the first two measurements are displayed below. The batch of third measurements is available in the file *prevention.sav*.

Data. The data matrix is given in the table below (Table 1-1). The variables are:

Group = indicates whether the participant had the treatment (1 = treatment, 2 = control)

Bdi1 = the total score on BDI before the treatment period (BDI pretest)

Bdi2 = the total score on BDI after the treatment period (BDI posttest)

Atq1 = the total score on ATQ before the treatment period (ATQ pretest)

Atq2 = the total score on ATG after the treatment period (ATQ posttest).

Table 1-1

Group	*Bdi1*	*Bdi2*	*Atq1*	*Atq2*
1	22	4	80	40
1	13	3	57	38
1	15	4	89	47
1	19		70	53
1	23	8	79	56
1	12	3	68	38
1	15	7	63	55
1	18	22	77	104
1	14	3	70	47
1	12	1	91	32
1	16	7	50	40
1	21	3	69	40
1	23	11	117	80
1	11	7	72	56
1	16	21	71	73
1	20	33	95	95
1	13	6	50	47
1	16	18	103	94
1	14	10	52	57
1	15	14	78	70
1	16	8	99	73
1	12	7	68	48
1	11	1	80	35
1	15	6	47	49
1	11	6	58	45
1	11	6	55	51
1	14	14	93	73
1	12	10	54	44
1	12	9	66	49
1	14	11	59	54
1	21	13	67	45
1	17	27	67	72
1	12	14	55	79
1	12	8	70	61
1	21	19	110	91
1	21	9	66	70
1	13	3	56	39
1	25	4	92	36
1	14	7	61	45
1	14	6	84	65

1	19	19	103	108
1	22	14	99	78
1	19	22	97	90
1	20	14	91	69
1	19	25	85	87
1	11	5	57	57
1	12	15	63	68
1	10	19	65	77
1	19	16	94	90
2	11	5	69	52
2	11	5	58	52
2	13	0	43	33
2	12	5	60	51
2	22	17	58	61
2	12	3	44	32
2	10	13	53	53
2	10	5	36	43
2	12	5	47	39
2	11	9	67	57
2	18	3	79	56
2	14	15	61	58
2	10	10	58	58
2	16	0	66	84
2	15	16	55	66
2	10	13	46	59
2	16	15	89	92
2	11	4	58	31
2	13	12	50	45
2	16	13	82	62
2	11	0	47	38
2	22	22	87	100
2	16	14	64	58
2	13	9	67	62
2	15	11	84	50
2	10	6	70	62
2	19	8	99	83
2	11	12	64	59
2	14	12	85	81
2	11	5	38	41
2	12	13	69	63
2	20	8	88	80
2	10	10	50	51

2	12	6	58	55
2	15	9	64	68
2	13	11	58	67
2	15	8	53	61
2	18	26	72	72
2	16	12	43	66
2	11	7	87	41
2	20	16	70	59
2	10	16	35	38

1.2 Example 2: Education (arithmetic lesson)

Reference of the study. This example is based on Ellis and Van de Veerdonk (2016).

Goal of the study. This study evaluated two different methods for teaching long division to students in intermediate vocational education.

Subjects. Participants were students of a school in the South-west of the Netherlands who were supposed to refresh their long-division skills.

Intervention. All participants followed a classroom lesson of one hour. Some classes were taught the classical long division algorithm. Other classes were taught the newer chunking method.

Measures. The proficiency at the long division was assessed before and after the lesson by two different though similar tests with 17 items each. The number of correct answers was converted into a grade between 1 and 10.

Data. The data matrix is given in the table below (Table 1-2). The variables are

Algorithm = indicates whether the student learned the chunking or the classical long division

GradeBefore = grade before the lesson (arithmetic pretest)

GradeAfter = grade after the lesson (arithmetic posttest)

Table 1-2

Person	*Algorithm*	*GradeBefore*	*GradeAfter*
1	chunk	5.8	7.4
2	chunk	6.3	5.8
3	chunk	8.4	7.9
4	chunk	5.2	6.8
5	chunk	6.8	6.8
6	chunk	8.4	7.9
7	chunk	6.8	6.8
8	chunk	6.8	8.9
9	chunk	7.9	7.4
10	chunk	3.1	1.5
11	chunk	7.4	4.7
12	chunk	8.9	5.8
13	chunk	8.4	6.3
14	chunk	8.9	7.4
15	chunk	5.2	6.3
16	chunk	4.7	5.2
17	chunk	4.7	5.2
18	chunk	5.2	6.8
19	classical	4.7	5.2
20	classical	5.2	7.9
21	classical	7.9	4.7
22	classical	8.4	6.8
23	classical	6.8	8.4
24	classical	7.4	6.3
25	classical	8.4	4.7
26	classical	6.3	6.3
27	classical	5.2	5.2
28	classical	7.4	4.7
29	classical	7.9	5.8

1.3 Example 3: Human-computer interaction (mind reading)

Reference of the study. This example is based on the study of Ryan et al. (2011).

Goal of the study. This study evaluated whether a predictive spelling program can improve the performance of a brain-computer interface (BCI) in typing sentences. Some diseases cause patients to become completely paralysed (locked-in syndrome), which makes communication extremely difficult. The BCI2000 software uses electrophysiological brain responses (EEG) to provide alphanumeric character selection by which patients can communicate. It is based on the P300, which is a positive part of the EEG response that occurs approximately 300 ms after presentation of a meaningful stimulus.

Subjects. Participants were 24 able-bodied undergraduate university students.

Intervention. All participants completed two sessions on separate days. In each session they had to write a sentence of 58 characters with their EEG, using the BCI2000 software. In one session the session (PS), the BCI2000 software was extended with a predictive spelling application. In the other session (NS), the BCI2000 software was used without predictive spelling. The order of the sessions was counterbalanced. This means that half of the subjects had the sessions in the order PS-NS, and the other half of the subjects had the sessions in the order NS-PS, and this order was assigned randomly to the subjects.

Measures. In this example, we will use some but not all of the measures of Table 1 and 2 of Ryan et al. (2011). The first measure is accuracy, which is the number of correct selections divided by the total number of selections in the session. The second measure is the time needed to complete the sentence, measured in minutes.

Data. The data matrix is given in the table below (Table 1-3). The variables are

PS Acc = accuracy in the PS condition

NS Acc = accuracy in the NS condition

PS Comp = time (minutes) to complete the sentence in the PS condition

NS Comp = time (minutes) to complete the sentence in the NS condition

Table 1-3 *Part of data of Ryan et al. (2011). Reprinted by permission of the publisher (Taylor & Francis Ltd, http://www.tandfonline.com).*

Subject	*PS Acc*	*NS Acc*	*PS Comp*	*NS Comp*
1	96.88	95.31	7.80	12.70
2	88.89	87.50	9.00	17.90
3	70.00	88.16	24.00	22.70
4	79.59	89.86	10.92	17.15
5	91.89	92.65	11.00	23.58
6	87.18	95.31	8.67	15.90
7	91.67	100.00	8.90	14.40
8	81.13	87.50	14.47	16.10
9	80.95	70.83	10.40	23.90
10	82.35	98.33	10.10	11.90

11	80.00	91.18	19.15	23.70
12	77.59	82.50	20.20	27.90
13	82.22	93.94	11.25	16.40
14	94.29	77.17	8.75	25.20
15	91.18	95.31	9.25	17.50
16	94.29	85.25	10.40	18.20
17	72.50	77.23	17.75	35.25
18	91.89	100.00	7.30	11.50
19	100.00	100.00	7.65	14.40
20	96.88	91.18	8.70	18.60
21	86.67	91.43	13.40	24.45
22	57.58	83.67	16.95	29.30
23	67.07	96.77	22.45	21.60
24	94.44	84.15	9.80	24.50

1.4 Example 4: Criminology (reconviction)

Reference of the study. This example is based on Killias, Gilliéron, Villard, and Poglia (2010).

Goal of the study. To compare the effect of imprisonment versus community service on re-offending.

Subjects. The participants were 141 defendants who had been sentenced to a short custodial sentence in the Swiss Canton of Vaud between 1993 and 1995. However, some subjects were excluded from the analysis because they had died, emigrated, had been removed from the programme, and for various other reasons.

Intervention. Subjects were randomly assigned to either community service or prison. Studies like this, in which real sentences are randomly assigned, are extremely exceptional.

Measures. It was recorded whether the subject has had at least one new conviction after 11 years.

Data. The counts are presented in Table 1 of Killias et al. (2010), and in the text pertaining to that table. In the custodial group, 22 out of 38 subjects had a new conviction after 11 years. In the community service group, 41 out of 78 subjects had a new conviction after 11 years.

1.5 Example 5: Developmental psychology (bullying)

Reference of the study. Unknown. The source of these data were developmental psychologists of the Radboud University.

Goal of the study. To investigate the effect of bullying on social isolation.

Subjects. 42 schoolchildren.

Measures. The children were questioned as to one another's social positions within the group. On the basis of this, the children were categorised as either bully, victim or non-involved. Each child responded to a questionnaire in order to measure social isolation. A high score on this variable means that the child was experiencing a lot of social isolation.

Data.

Table 1-4

Position	*Social isolation*
non-involved	1.75
non-involved	0.25
non-involved	0.25
non-involved	1.25
non-involved	0.75
non-involved	0.5
non-involved	0.25
non-involved	1
non-involved	0.75
non-involved	1
non-involved	1.25
non-involved	0.25
non-involved	1.5
non-involved	0.5
non-involved	0.25
non-involved	0.25
non-involved	0.25
non-involved	0.5
non-involved	1.75
non-involved	0.25
bully	0.75
bully	0.25
bully	0.75
bully	0.5

bully	0.5
bully	0.5
bully	0.75
bully	0.25
bully	0.25
victim	1
victim	2
victim	0.75
victim	3.25
victim	0.5
victim	2.25
victim	3
victim	1
victim	1.75
victim	1.75
victim	0.25
victim	0.25
victim	1.75

1.6 Example 6: Social psychology (food consumption)

Reference of the study. This example is based on Van de Veer, Van Herpen, and Van Trijp (2015).

Goal of the study. To investigate the effect of outer appearance focus on food consumption.

Subjects. The study does not specify how the participants were initially sampled. Participants who did not follow the instructions and participants who indicated a strong dislike for M&Ms were excluded from the study. Also, a participant who was an outlier in mean M&M consumption was excluded. There remained 107 subjects.

Interventions. In the appearance focus condition, a mirror was positioned so that participants could view their face and upper body part. In a previous study it was found that this increases attention to outward appearance of the body. In the condition of no appearance focus, the mirrors were turned back to front. Participants were served a 300-ml milkshake that contained either 534 or 215 kcal. So, in total there were two focus conditions (appearance focus vs no appearance focus) and two conditions as to the milkshake's caloric content (low vs high). The design was crossed, which means that all $2 \times 2 = 4$ combinations of focus and caloric content were used. Participants