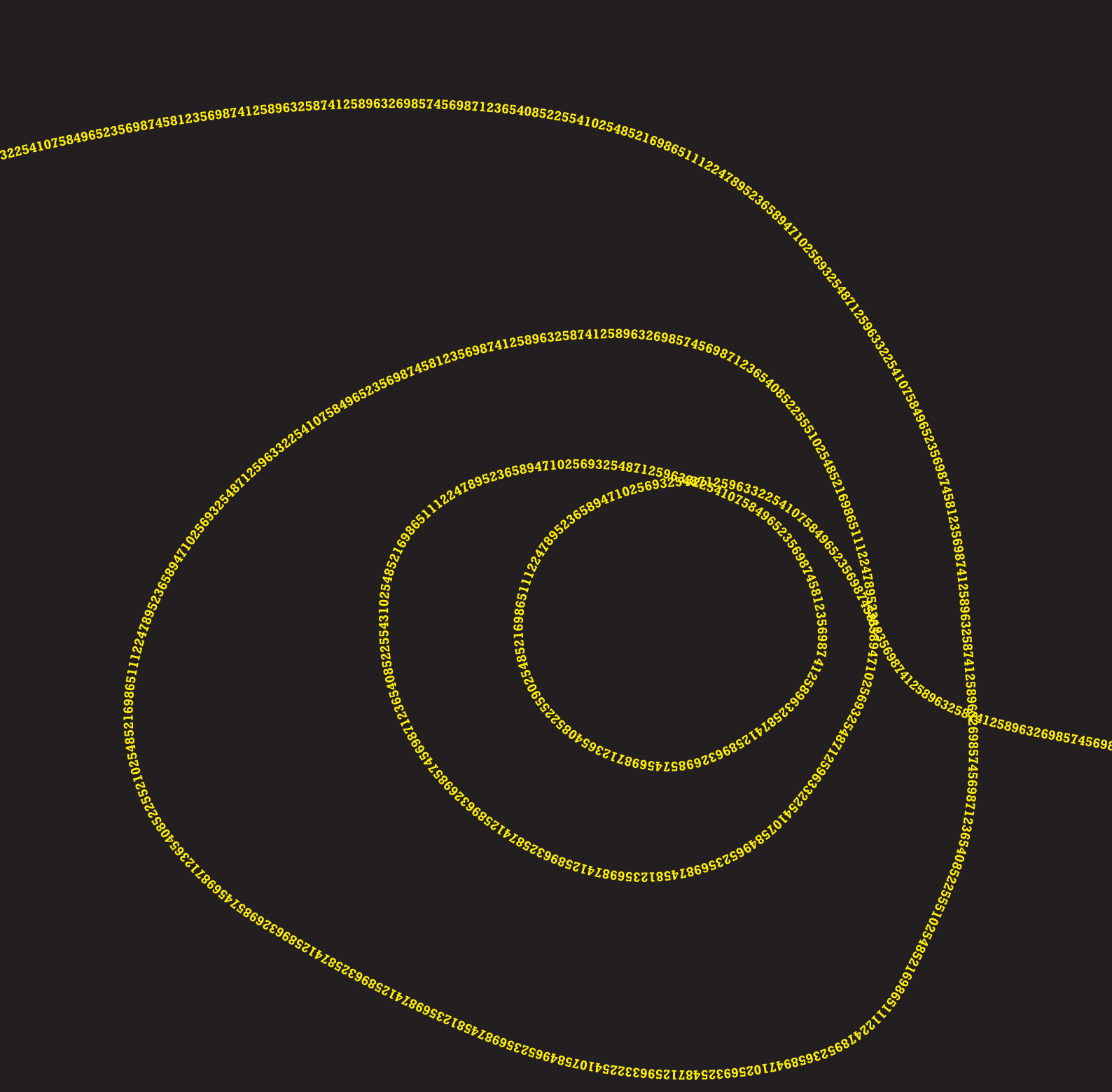
SPSS STATISTICS

A PRACTICAL GUIDE | 5TH EDITION

KELLIE **BENNETT** BRODY **HERITAGE** PETER **ALLEN**

SPSS STATISTICS

A PRACTICAL GUIDE | 5TH EDITION



SPSS STATISTICS

A PRACTICAL GUIDE | 5TH EDITION

KELLIE BENNETT BRODY HERITAGE PETER ALLEN

SPSS Statistics: A Practical Guide
5th Edition
Kellie Bennett
Brody Heritage
Peter Allen

Portfolio manager: Fiona Hammond
Product manager: Michelle Aarons
Content developer: Lynley Bidlake/Eleanor Yeoell
Project editor: Alex Chambers
Cover designer: Cengage Creative Studio
Editor: Jennifer Bulter
Indexer: Julie King
Art direction: Linda Davidson
Typeset by KnowledgeWorks Global Ltd.

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

This 5th edition published in 2023.

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email **aust.permissions@cengage.com**

National Library of Australia Cataloguing-in-Publication Data
ISBN: 9780170460163
A catalogue record for this book is available from the National Library of Australia.

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit **cengage.com.au**

Printed in China by 1010 Printing International Limited.
1 2 3 4 5 6 7 26 25 24 23 22

BRIEF CONTENTS

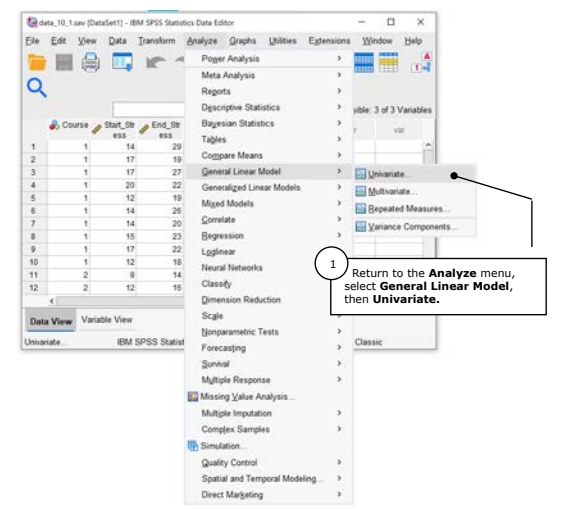
CONTENTS

Guide to the text

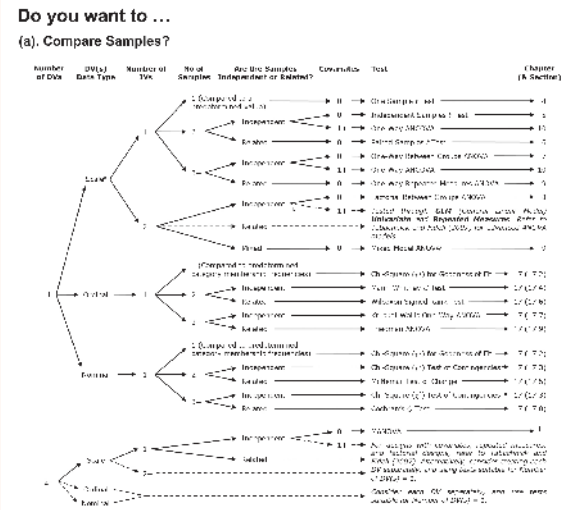
As you read this text you will find a number of features in every chapter to enhance your study of SPSS Statistics and help you understand how the theory is applied in the real world.

FEATURES WITHIN CHAPTERS

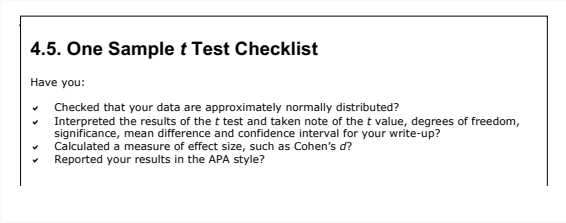
Step-by-step instructions and **annotated screenshots** clearly and visually explain each procedure to help students easily understand how to use SPSS Statistics version 28.



A **Decision tree** printed inside the front cover quickly guides you to appropriate procedures for your circumstances.



Checklists at the end of each analysis chapter remind you of key concerns and the steps to complete.



4.5. One Sample *t* Test Checklist

Have you:

- Checked that your data are approximately normally distributed?
- Interpreted the results of the *t* test and taken note of the *t* value, degrees of freedom, significance, mean difference and confidence interval for your write-up?
- Calculated a measure of effect size, such as Cohen's *d*?
- Reported your results in the APA style?

ICONS

Ask your instructor for the data and syntax sets used in the examples. Look out for the Try it out icons and follow the step-by-step guide exactly.

The AKA icons indicate where key terms could be described differently in other resources.

Use the Tip icons for handy hints to complete the step.

Follow the **LINK icons** to revisit useful topics from other chapters as you go along

Try It Out:
This is data file **data_4_1.sav**.

AKA:
Single sample *t* test.

Tip:
Report the absolute value of *d*, rather than a negative value.

Link:
Setting up a data file is illustrated in greater detail in section 1.2.

Guide to the online resources

FOR THE INSTRUCTOR

Cengage is pleased to provide you with a selection of resources that will help you to prepare your lectures and assessments, when you choose this textbook for your course.
Log in or request an account to access instructor resources at **au.cengage.com/instructor/account** for Australia or **nz.cengage.com/instructor/account** for New Zealand.

DATASETS AND SYNTAX SETS

All the **datasets and syntax sets** used in the examples.

INSTRUCTOR'S MANUAL

The **Instructor's manual** includes:

- Author's guide to the resources
- Practice exercises with solutions
- Practice exercise worksheets and matching datasets
- Revision quizzes

POWERPOINT™ PRESENTATIONS

Use the chapter-by-chapter **PowerPoint slides** to enhance your lecture presentations and handouts by reinforcing the key principles of your subject.

ARTWORK FROM THE TEXT

Add the digital screenshots into your course management system, use them in student handouts, or copy them into your lecture presentations.

PREFACE

IBM SPSS Statistics is a flexible set of data analytic tools used throughout many disciplines. Its roots can be traced back as far as 1968, when a small group of Stanford University doctoral students began developing the *Statistical Package for the Social Sciences* in response to their own need for a software system that would allow them to efficiently analyse the large amounts of data they were amassing at that time.

This program quickly grew beyond the cloistered confines of Stanford, with the publication of the first user manual in 1970. *SPSS Statistics: A Practical Guide* continues this long tradition of opening statistical analysis up to students and early career researchers from a wide range of applied and academic disciplines. We hope you find it useful!

About IBM SPSS Statistics Version 28

IBM released *IBM SPSS Statistics Version 28* in late 2021. SPSS Statistics is available for a number of operating systems, including Microsoft Windows, Macintosh OSX, and several Linux distributions.

For more information about SPSS Statistics, visit https://www.ibm.com/au-en/analytics/spss-statistics-software.

About this book

With *SPSS Statistics: A Practical Guide* we aim to introduce readers to a range of commonly used statistical procedures that can be performed with SPSS Statistics, and are typically included in the curricula of undergraduate applied statistics and research methodology units.

Our approach is unashamedly practical, and highly visual. We take a hands-on approach to our subject matter, and work through each procedure in an illustrated, step-by-step fashion. Beyond the necessary focus on "doing", we emphasise interpretation and reporting throughout this text, which contains hundreds of tables of tightly annotated SPSS Statistics output, and dozens of examples of how specific research findings can be communicated clearly and concisely.

Other prominent features of this text include:

- Illustrated examples of statistically "significant" and "non-significant" findings, recognising that real-life data does not always support our hypotheses.
- Guidelines for calculating and interpreting effect sizes for most of the inferential procedures discussed in this book.
- Power calculations and Baysian analyses are discussed.
- An ongoing emphasis on assumption testing, with recommendations for dealing with violated assumptions.
- An extensive section on non-parametric procedures.
- Online resources including datasets and syntax files, which are available at http://login.cengagebrain.com.au.

The screen captures used throughout the book were taken with SPSS Statistics version 28 for Windows. If you're using an older (or newer) version of SPSS Statistics, or SPSS Statistics for Macintosh OSX or Linux, you may notice small differences between our captures and the windows and dialogue boxes on your own screen. In virtually all instances, these differences will be slight, and should not hinder your ability to follow our worked examples. However, if you do notice any substantial differences between your version of SPSS Statistics and ours, please let us know, and we'll be sure to make a note of it in the next edition of this text.

We do not anticipate that you will read this text from cover to cover. Rather, we hope you'll be able to pick it up and identify quickly the sections you need to 'get the job done'. To make this task easier, we've divided our content into 18 conceptually distinct chapters.

Chapters 1 to 3 introduce new users to the SPSS Statistics interface, and to some of the many ways SPSS Statistics can be used to manipulate, summarise and display data. Chapters 4 to 17 are dedicated to specific inferential procedures, including:

- t tests (one sample; independent samples; and paired samples).
- Analysis of variance (one-way and factorial; between groups, repeated measures and mixed).
- Analysis of covariance.
- Multivariate analysis of variance.
- Bivariate and partial correlation.
- Multiple regression (both standard and hierarchical).
- Logistic regression (focusing on binary logistic regression).
- Factor analysis.
- Reliability analysis.
- Non-parametric alternatives (including chi-square tests for goodness of fit and contingencies; Mann-Whitney's U; McNemar's test of change; Wilcoxon's signed rank test; the Kruskal-Wallis ANOVA; the Friedman ANOVA; Cramer's V; Spearman's rho; and Kendall's tau-b).

Within each of these chapters, we outline the purpose of the test(s), and illustrate the types of research questions they can be used to address. We then step the reader through one or more illustrated examples, from initial assumption testing through to follow-up analyses and effect-size estimation. Each example then concludes with an annotated APA (American Psychological Association) style results section, demonstrating exactly how research findings can be clearly communicated in reports, assignments and poster presentations. All the examples and research findings discussed in this text are for illustrative purposes only. The datasets have been created by the authors and are not based on actual research studies.

Finally, Chapter 18 looks at the use of SPSS Statistics command syntax, and the flexibility and efficiencies it can offer more advanced users.

Some final comments

This is not a statistics textbook, and we are not mathematicians. We use statistics as a means to an end; as a tool for managing and making sense of the data we collect as part of our efforts – through our research – to better understand the world around us. We're assuming that most of our readers feel the same way, and are using SPSS Statistics to take some of the time-intensive number-crunching out of data analysis (thus freeing us all up to do more important things, like interpreting our findings and communicating them to our colleagues and beyond).

Having said that, we do assume that you have a basic understanding of applied statistics, as well as issues of research design more broadly. If not (or if you're just feeling a bit rusty), we recommend using our text alongside books that do justice to these important issues.

There are many such books available. For introductory level statistics, we recommend *Gravetter, Wallnau & Forzano's* (2018) Essentials of Statistics for the Behavioral Sciences and Howell's (2017) *Fundamental Statistics for the Behavioral Sciences.*

Happy analysing!

Kellie Bennett
Division of Psychiatry, Medical School University of Western Australia

Brody Heritage
Discipline of Psychology, Murdoch University
Telethon Kids Institute, University of Western Australia

Peter Allen
School of Experimental Psychology University of Bristol

February 2022

ABOUT THE AUTHORS

Dr Kellie Bennett is in the Medical School at the University of Western Australia. Kellie teaches in the medical and undergraduate science programmes, supervises postgraduate students and has provided statistical advice for a number of large research projects. Kellie's main research interests include psychological aspects of mental health, medical communication, developmental consequences of substance use, and related statistical analysis and research methodology.

Dr Brody Heritage was in the Discipline of Psychology, Murdoch University, and the Telethon Kids Institute, University of Western Australia, at varying parts of the authorship of this edition. Brody thoroughly enjoys research involving quantitative scale development and refinement through psychometric studies. He also has an interest positive psychology research, specifically in how it can be relevant for younger people in school, and how this can help them work through the challenges they encounter.

Peter Allen is a Senior Lecturer in the School of Psychological Science at the University of Bristol in the United Kingdom. Prior to this he was a Lecturer in the School of Psychology and Speech Pathology at Curtin University. Peter has over 15 years of experience teaching research methods and statistics to undergraduate students. His research interests include evidence-based learning and teaching, with a particular emphasis on the development of statistical decision-making skills.

ACKNOWLEDGEMENTS

The publisher and authors would like to thank the academics who reviewed original chapters of this text and provided feedback, including Lynne Roberts and Nick Barrett, Curtin University; and all the students and colleagues who provided feedback on the previous version of this text including Harold Hill, University of Wollongong; Einar Thorsteinsson, University of New England and others.

PROVIDING FEEDBACK

Your feedback is important, and helps us continue developing and improving on subsequent editions of this text. Please do let us know if you notice any errors or omissions in this text, or if there are SPSS Statistics procedures or pedagogic features you would like to see included in future editions. You can contact us at anz.customerservice@cengage.com.

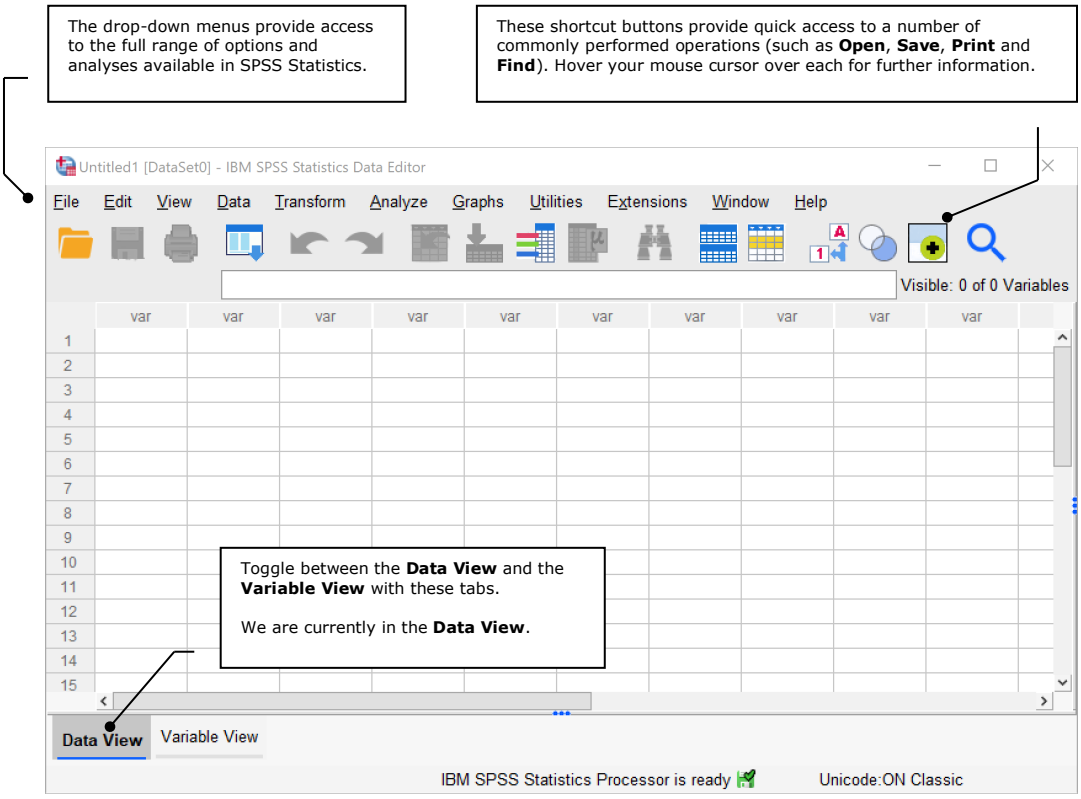
Chapter 1: Getting Started With SPSS Statistics

Chapter Overview

1.1. Introduction

This chapter has two purposes: (a) to introduce the **SPSS Statistics Data Editor**; and (b) to step you through the process of setting up a simple SPSS Statistics data file.

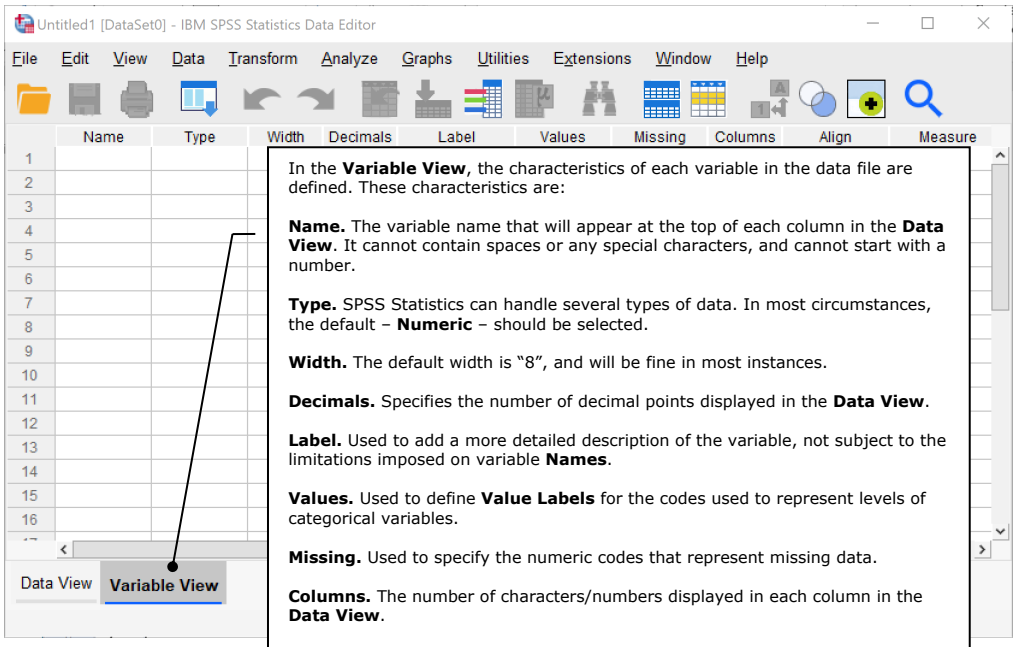
1.1.1. Data View



Link:
The **Data Editor** is just one component of the SPSS Statistics working environment. Other components discussed in this book include:

- **Viewer** is used in chapter 3 (section 3.2.1.3)
- **Pivot Table Editor** is used in chapter 3 (section 3.2.1.3)
- **Chart Editor** is used in chapter 3 (section 3.4.1.3)
- **Syntax Editor** is used in chapter 18 (section 18.2.2)

1.1.2. Variable View



In the **Variable View**, the characteristics of each variable in the data file are defined. These characteristics are:

Name. The variable name that will appear at the top of each column in the **Data View**. It cannot contain spaces or any special characters, and cannot start with a number.

Type. SPSS Statistics can handle several types of data. In most circumstances, the default – **Numeric** – should be selected.

Width. The default width is "8", and will be fine in most instances.

Decimals. Specifies the number of decimal points displayed in the **Data View**.

Label. Used to add a more detailed description of the variable, not subject to the limitations imposed on variable **Names**.

Values. Used to define **Value Labels** for the codes used to represent levels of categorical variables.

Missing. Used to specify the numeric codes that represent missing data.

Columns. The number of characters/numbers displayed in each column in the **Data View**.

Align. Used to specify whether the variable is **Left**, **Right** or **Center** aligned in the **Data View**.

Measure. Used to specify whether the variable is **Scale, Ordinal** or **Nominal**. SPSS Statistics does not discriminate between interval and ratio data. It refers to them both as **Scale** data.

Role. Used to set predefined roles for variables, so they can be automatically sorted into appropriate lists in selected SPSS Statistics dialogues. As many dialogues do not yet support roles, the default option (**Input**) will usually be fine. See the Tip in Section 1.2 for more information.

ⓘ ***Tip:***
Work in rows in the **Variable View**, where each row defines one variable in your SPSS Statistics data file.

ⓘ ***Tip:***
Interval and ratio data (which are both referred to as **Scale** data in SPSS Statistics) have the properties we tend to associate with "real numbers".

Ordinal data can be ranked (i.e., an ordinal scale has magnitude), but they lack any other numeric properties.

Nominal data are categorical, and the values we assign to levels of a nominal variable are nothing more than shorthand labels. They have no true numeric properties.

1.2. Creating a Data File

To illustrate the process of creating and setting up a data file, we developed a simple survey, and asked 15 students in the Health Sciences café to complete it. We've reproduced four of the completed surveys in full, and the data that were collected with the remaining 11 are summarised in Table 1.1.

❑ ***Try It Out:*** This is data file **data_1_1.sav**.

Participant 1

A Quick Student Satisfaction Survey!

1. Gender: Male ☑ Female ☐ *(Please tick)*

2. Age: 17

3. Course of Study: Speech therapy

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate the extent to which you agree with each of the following statements:

4. I am enjoying my course.

SD 1 2 3 4 (4) *SA*

5. It is easy to get good grades in my course.

SD 1 2 3 4 (5) *SA*

Thanks for completing our survey!

Participant 2

A Quick Student Satisfaction Survey!

1. Gender: Male ☑ Female ☐ *(Please tick)*

2. Age: _____

3. Course of Study: Physiotherapy

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate the extent to which you agree with each of the following statements:

4. I am enjoying my course.

SD 1 2 3 (4) *5 SA*

5. It is easy to get good grades in my course.

SD 1 2 (3) *4 5 SA*

Thanks for completing our survey!

Participant 3

A Quick Student Satisfaction Survey!

1. Gender: Male ☐ Female ☑ *(Please tick)*

2. Age: 18

3. Course of Study: OT

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate the extent to which you agree with each of the following statements:

4. I am enjoying my course.

SD 1 2 3 4 (5) *SA*

5. It is easy to get good grades in my course.

SD 1 2 3 (4) *5 SA*

Thanks for completing our survey!

Participant 4

A Quick Student Satisfaction Survey!

1. Gender: Male ☒ Female ☐ *(Please tick)*

2. Age: 17

3. Course of Study: PHYSIO.

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate the extent to which you agree with each of the following statements:

4. I am enjoying my course.

SD 1 2 3 (4) *5 SA*

5. It is easy to get good grades in my course.

SD 1 2 3 (4) *5 SA*

Thanks for completing our survey!

Table 1.1

Data Collected From Participants 5-15 With "A Quick Student Satisfaction Survey"

ID	Q1	Q2	Q3	Q4	Q5
5	Female	20	Physiotherapy	4	4
6	Female	21	Psychology	3	3
7	Female	19	Speech Therapy	3	2
8	Male	18	Speech Therapy	2	1
9	Male	19	Nursing	3	3
10	Female	21	Public Health	5	4
11	Female	24	Occupational Therapy	4	3
12	Female	17	Occupational Therapy	1	1
13	Female	38	Nursing	3	3
14	Male	19	Occupational Therapy	4	5
15	Female	18	Occupational Therapy	5	4

ⓘ ***Tip:***
Some dialogues in SPSS Statistics 28 support **Roles**.

When you use one of these dialogues, your variables will be automatically sorted into lists according to their defined roles.

There are six possible roles to select from:

Input. Should be selected for predictor or independent variables.

Target. Should be selected for output or dependent variables.

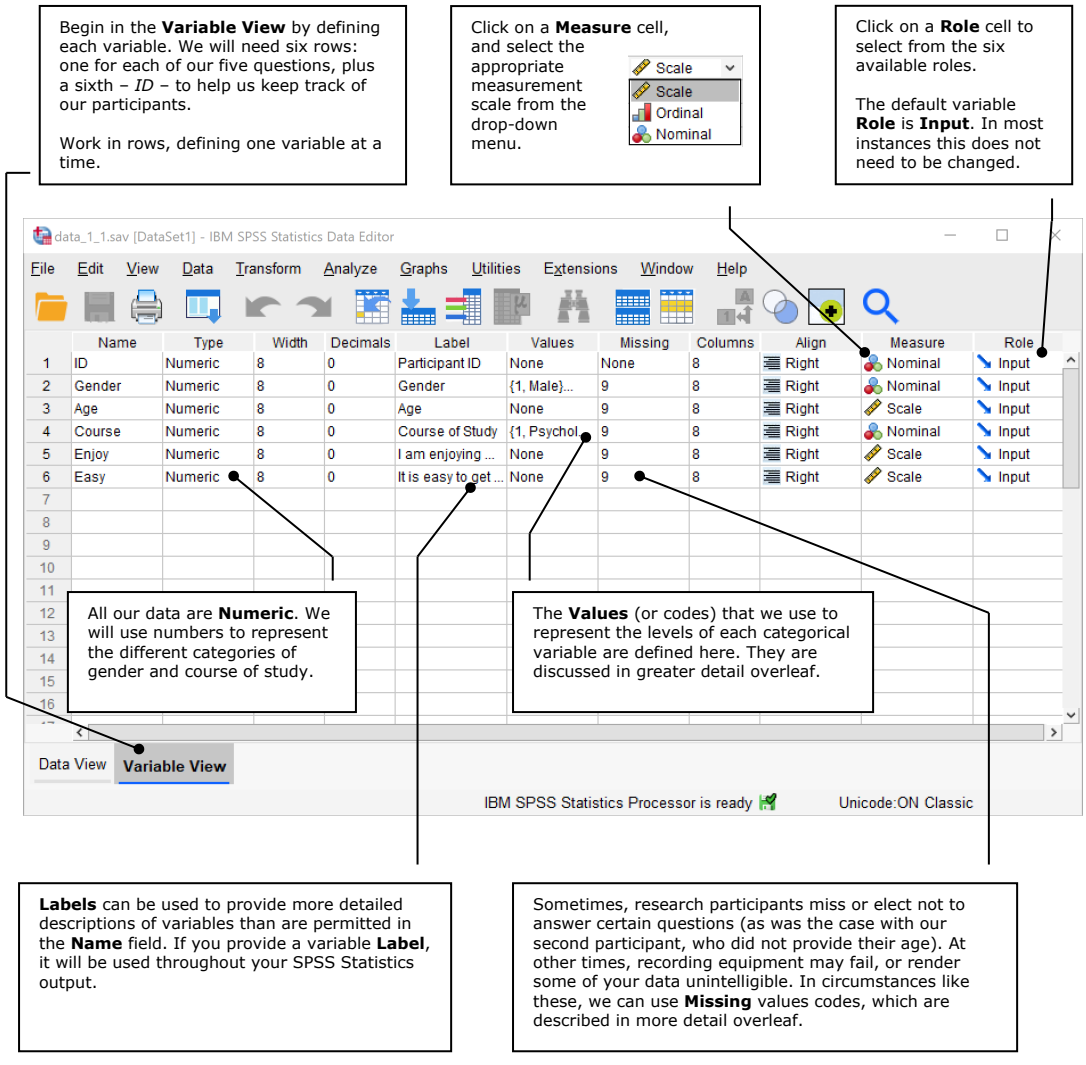
Both. Should be selected for variables that will be used as both predictor and output variables.

None. Should be selected for variables that you do not want to assign a specific role to.

Partition. Used to partition the data file into separate samples.

Split. Included for compatibility with **SPSS Modeler**.

As most SPSS Statistics dialogues currently ignore variable **Roles**, you can too!



Value Labels

In SPSS Statistics, we often use numeric **Values** (or codes) to represent levels of categorical (or **Nominal**) variables, like gender or course of study.

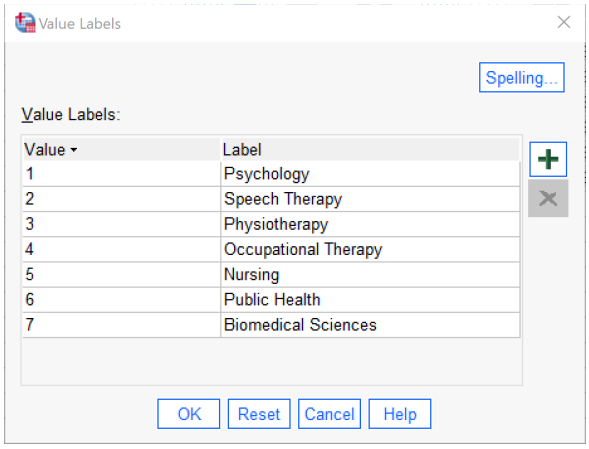
We define these codes in the **Value Labels** dialogue.

To open the **Value Labels** dialogue, select a **Values** cell in the **Variable View**, then click ....

Type the number used to represent the first category (or level) in the **Value** box, and the category name in the **Label** box.

Click the **Add** button to add them to the **Value Labels** list.

Repeat this process for the remaining categories, then click **OK** to close the **Value Labels** dialogue.



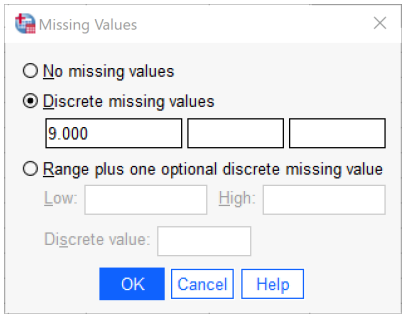
Here, we have used the values "1" through "7" to represent the seven courses that our student participants are enrolled in.

Missing Values

It is inevitable that, from time to time, you will be confronted with missing data. Participants sometimes miss questions, recording equipment fails, coffee spills smudge answers, and so on. We can record unfortunate situations like these with **Missing Values** codes.

At the simplest level, we can use one numeric code to represent all types of missing data. Or, we can discriminate between different sorts of missing data (e.g., questions that participants refused to answer, versus questions which were missed due to equipment failure) by specifying up to three unique **Missing Values** codes.

You can use any numeric code(s) to represent missing data, provided they are outside the range of your actual data. You can also use different codes on each variable, or the same code(s) throughout the entire data file.

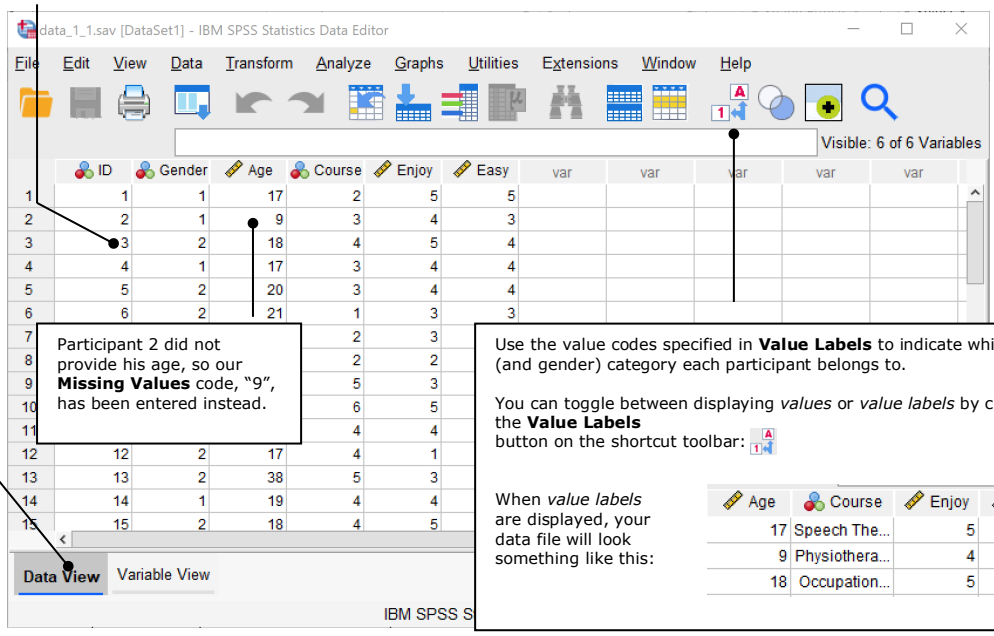


We've selected "9" as our **Missing Values** code because it can't be confused with "real" data on any of our research variables. (For example, none of our participants will have a gender of "9", or will have responded with "9" to any of our 5-point rating scales.)

The **Missing Values** dialogue is accessed by selecting a cell under **Missing** in the **Variable View**, then clicking the ... button. **Missing Values** codes must be specified separately for each variable.

Tip:
If you ever need to discriminate between more than three types of missing data, you can select **Range plus one optional discrete missing value** and enter a range of values you want SPSS Statistics to treat as "missing".

After defining each variable, you can begin typing in data. In the **Data View**, each row represents a case (e.g., a participant). For example, row 3 contains the data provided by participant 3, an 18-year-old Occupational Therapy student.



Participant 2 did not provide his age, so our **Missing Values** code, "9", has been entered instead.

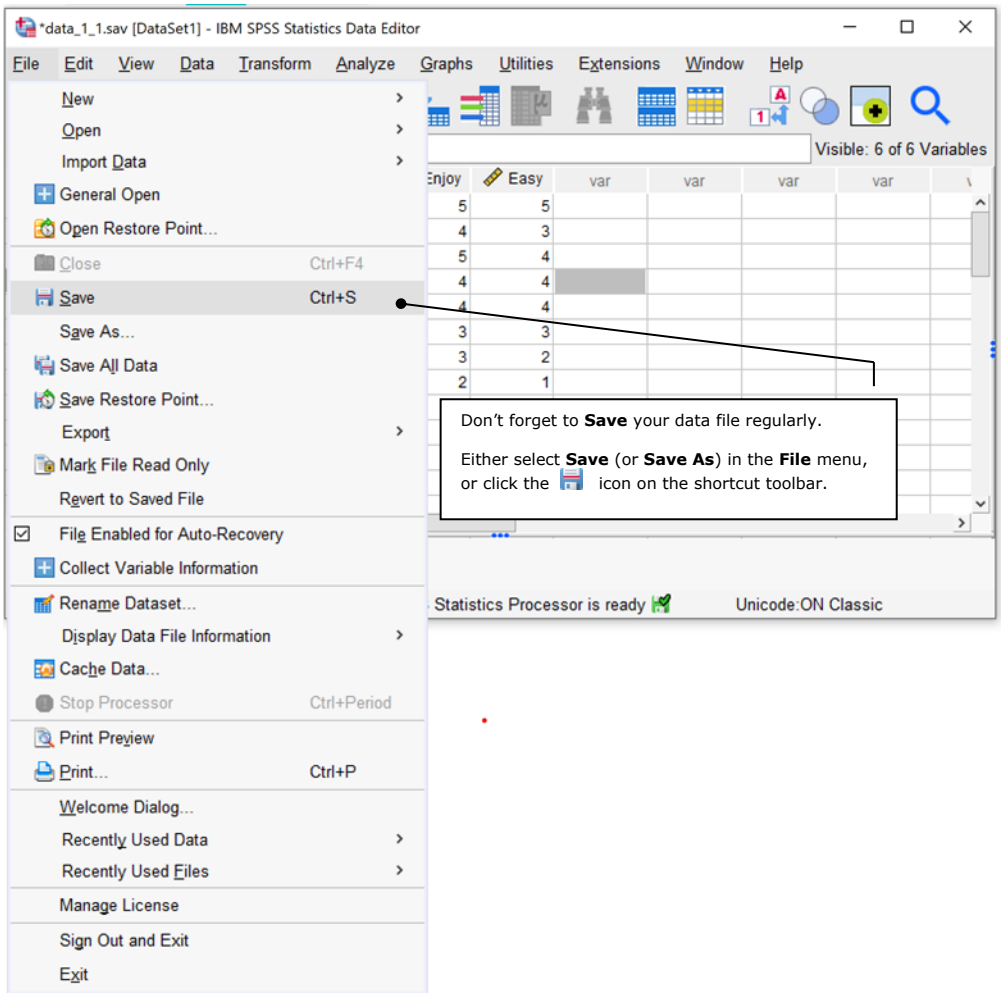
Use the value codes specified in **Value Labels** to indicate which course (and gender) category each participant belongs to.

You can toggle between displaying *values* or *value labels* by clicking the **Value Labels** button on the shortcut toolbar:

When *value labels* are displayed, your data file will look something like this:

Age	Course	Enjoy	Easy
17	Speech The...	5	5
9	Physiothera...	4	3
18	Occupation...	5	4

1.3. Conclusion



With the data saved, we can begin working with it, summarising it and displaying it. These are the topics covered in chapters 2 and 3.

Chapter 2: Working With Data

Chapter Overview

2.1. Introduction

In chapter 1 we described a short survey used to collect some demographic and course satisfaction data from 15 Health Sciences students. In the current chapter, we will continue using this data to illustrate how a data file can be manipulated in SPSS Statistics.

The data we collected are reproduced in Table 2.1.

Table 2.1

Data Collected From 15 Participants With "A Quick Student Satisfaction Survey"

❑ ***Try It Out:*** This is data file **data_2_1.sav**.

ID	Q1[a]	Q2	Q3[b]	Q4	Q5
1	Male	17	Speech Therapy	5	5
2	Male		Physiotherapy	4	3
3	Female	18	Occupational Therapy	5	4
4	Male	17	Physiotherapy	4	4
5	Female	20	Physiotherapy	4	4
6	Female	21	Psychology	3	3
7	Female	19	Speech Therapy	3	2
8	Male	18	Speech Therapy	2	1
9	Male	19	Nursing	3	3
10	Female	21	Public Health	5	4
11	Female	24	Occupational Therapy	4	3
12	Female	17	Occupational Therapy	1	1
13	Female	38	Nursing	3	3
14	Male	19	Occupational Therapy	4	5
15	Female	18	Occupational Therapy	5	4

Note. Q1 = Gender; Q2 = Age; Q3 = Course of study; Q4 = I am enjoying my course (from 1 = strongly disagree to 5 = strongly agree); Q5 = It is easy to get good marks in my course (from 1 = strongly disagree to 5 = strongly agree).

[a] Value labels for gender are 1 = male; 2 = female.

[b] Value labels for course of study are 1 = Psychology; 2 = Speech Therapy; 3 = Physiotherapy; 4 = Occupational Therapy; 5 = Nursing; 6 = Public Health; 7 = Biomedical Sciences.

Try It Out: Run these analyses with **syntax_2_1.sps**.

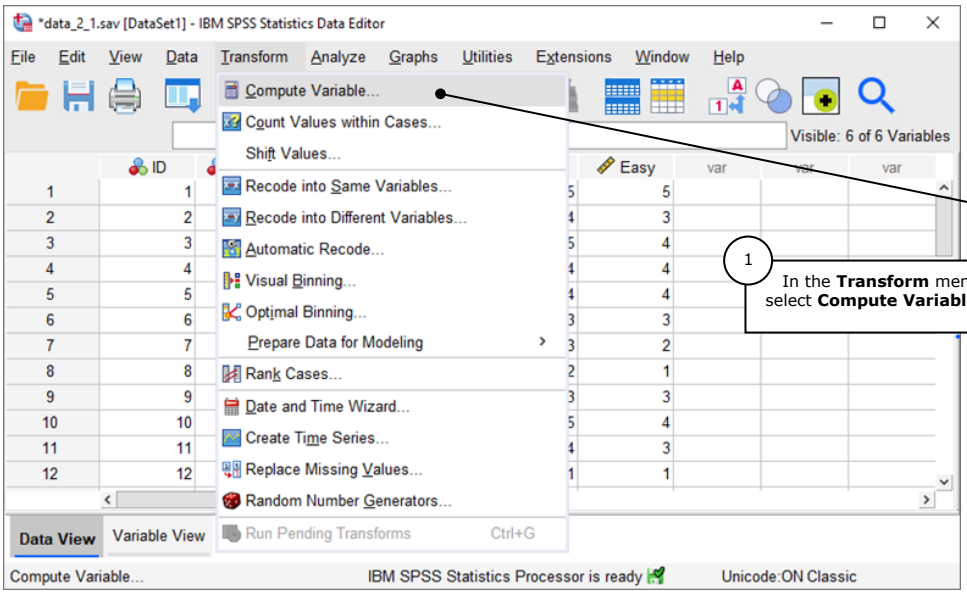
2.2. Compute

Compute Variable allows you to create a new variable from one or more existing variables. You can also use it to alter an existing variable.

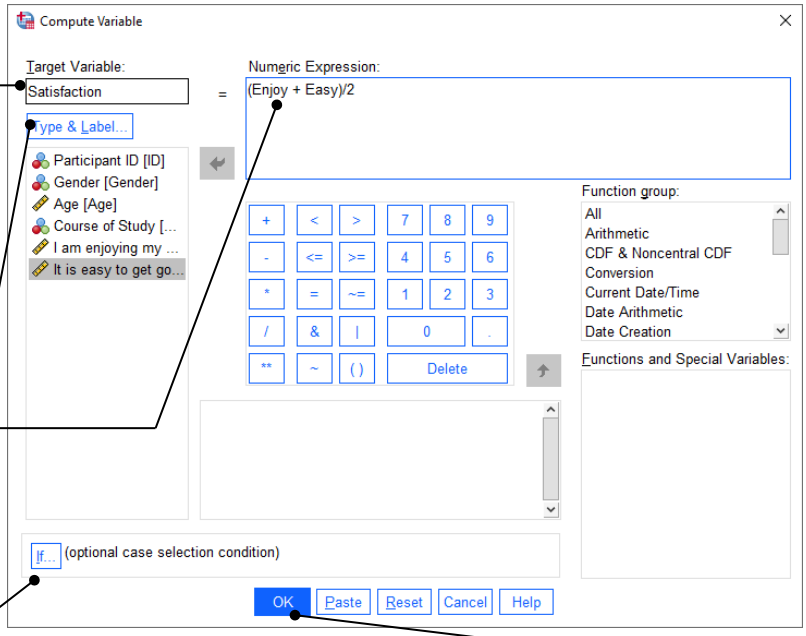
2.2.1. Illustrated Example of Summed Scale Scores

Most commonly, we use **Compute Variable** to sum or average participants' responses to questionnaire items. Here, we'll create a new variable – *course satisfaction* – by averaging our participants' answers to Q4 and Q5.

Link: **Compute Variable** is used in chapter 6 (section 6.3.2.2) to create a difference scores variable.



1 In the **Transform** menu, select **Compute Variable**.



2 Give the new variable a **Name** in the **Target Variable** box.

3 In the **Type & Label** dialogue, enter a more informative **Label** for the new variable (e.g., *course satisfaction*).

4 Build the **Numeric Expression** that will be used to calculate values on the new variable. Here:

(*Enjoy* + *Easy*) / 2

(*Enjoy* and *Easy* are the **Names** of the variables being averaged.)

Type directly into the **Numeric Expression** box, or build the expression using the keypad, arrow button and variable list.

5 Clicking the **If** button opens the **If Cases** dialogue, in which you can specify which cases the **Numeric Expression** should be applied to.

For example, the conditional statement "*Gender* = 1" would apply the "(*Enjoy* + *Easy*) / 2" expression to the male cases only.

For our purposes though, we will calculate a satisfaction score for every case in our data file.

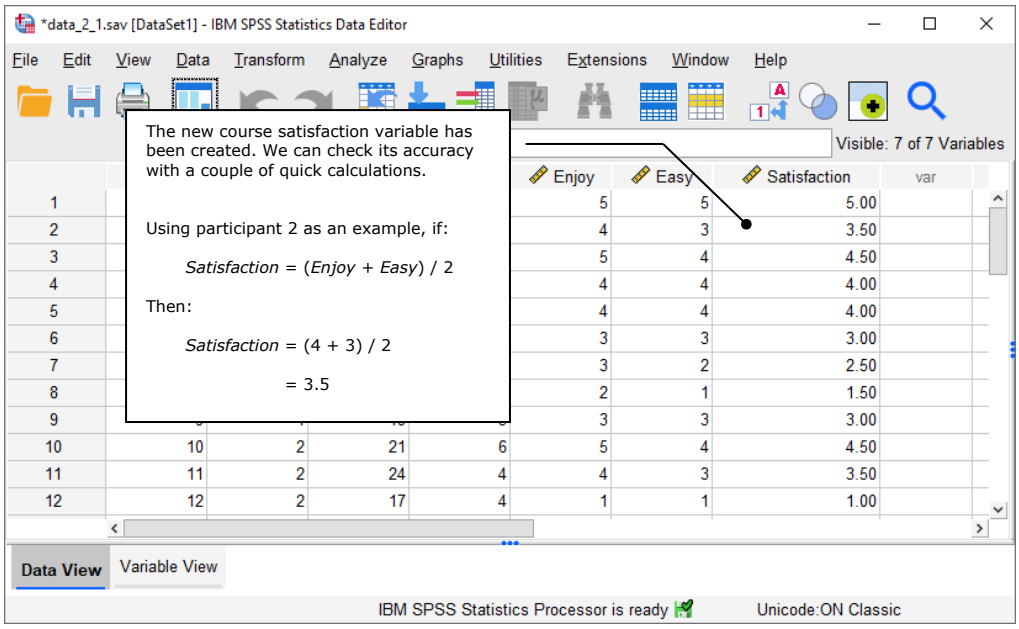
6 Click **OK** to create the new variable.

Tip: SPSS Statistics includes many functions that can be used to compute new variables.

To access these functions, select a **Function group** and then choose from the options available in the **Functions and Special Variables** list.

When a selection is made, the function's description is provided in the space to the right of the variable list.

If the selected function meets your requirements, it can be moved into the **Numeric Expression** box with the arrow button: ↑



2.3. Recode

With **Recode** you can change specific values or ranges of values on one or more variables. This feature can be used to:

a. *Collapse continuous variables into categories.*

We could collapse the *satisfaction* variable into two categories ("not satisfied" or "satisfied") by recoding scores of 3 or less as "1" (where 1 = not satisfied) and scores higher than 3 as "2" (where 2 = satisfied).

b. *Combine or merge values.*

We could merge the courses into two categories – those that require new enrollees to have studied high-school human biology, and those that don't.

c. *Reverse negatively scaled questionnaire items.*

Link:
Recode is used to reverse negatively scaled questionnaire items in chapter 16 (section 16.2.1.2).

Oftentimes, questionnaires contain both positively and negatively worded items. For example:

1. I am enjoying my course. *Strongly Disagree 1 2 3 4 5 Strongly Agree*
2. I hate my course. *Strongly Disagree 1 2 3 4 5 Strongly Agree*

On this questionnaire, a participant who answers "5" to the first question would likely answer "1" to the second question, and summing or averaging these responses would make very little sense. However, by reversing his or her response to the second question (by recoding 1 as 5, 2 as 4, and so on) we are able to calculate a meaningful total or average, which can then be used in subsequent analyses.

d. *Replace missing values.*

e. *Bring outliers and extreme scores closer to the rest of the distribution.*