

Using the SPSS training materials creation features of Stat-JR (1.0.6)

This documentation was written by William Browne, Chris Charlton and Liz Washbrook

Centre for Multilevel Modelling,
University of Bristol

November 2018

Using the SPSS training creation features of Stat-JR (1.0.6)

© 2018. William J. Browne, Christopher M.J. Charlton and Elizabeth Washbrook.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, for any purpose other than the owner's personal use, without the prior written permission of one of the copyright holders.

ISBN: To be confirmed

Printed in the United Kingdom

Funding acknowledgement

We are grateful to the British Academy for funding the work upon which this guide has been developed. We are also grateful to the ESRC for the financial support allowing us to create the underlying Stat-JR software.

Contents

Introduction	1
Linking data to a skeletal SPSS eBook	1
Reading and printing the eBook	4
Customisation	8
1. Exploring the example	8
2. Reproducing the supplied PISA PDF.....	11
3. Using your own dataset	13
Appendix – Background to the PISA datafile	15

Introduction

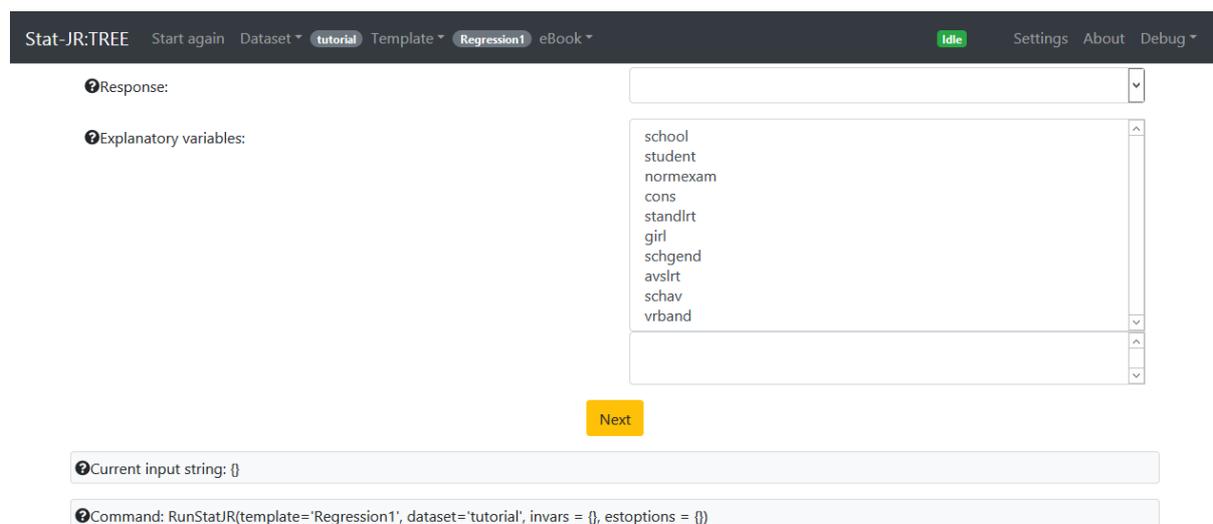
This document accompanies the new Stat-JR templates and eBooks that were created as part of a grant from the British Academy. This documentation also uses the OECD's PISA data (see <http://www.oecd.org/pisa/data/2015database/>) that has been extracted, processed and stored in .dta format. The new template, eBook and data files are available from <http://www.bristol.ac.uk/cmm/software/statjr/downloads/> along with instructions for extracting them. Details on how these files were created and what they contain can be found in the Appendix to this document.

The remainder of the document takes the reader through constructing SPSS training materials using Stat-JR. The document doesn't contain background information on Stat-JR so for this we advise the reader to look at the extensive documentation at <http://www.bristol.ac.uk/cmm/software/statjr/manuals/> in particular the Quick Start guide and Beginner's guide.

Linking data to a skeletal SPSS eBook

The TREE module within Stat-JR allows the user to interactively run models, manipulate data and perform data summarisation. The outputs from these can be combined into an eBook form via a word-processing like interface, which will be the focus of this document.

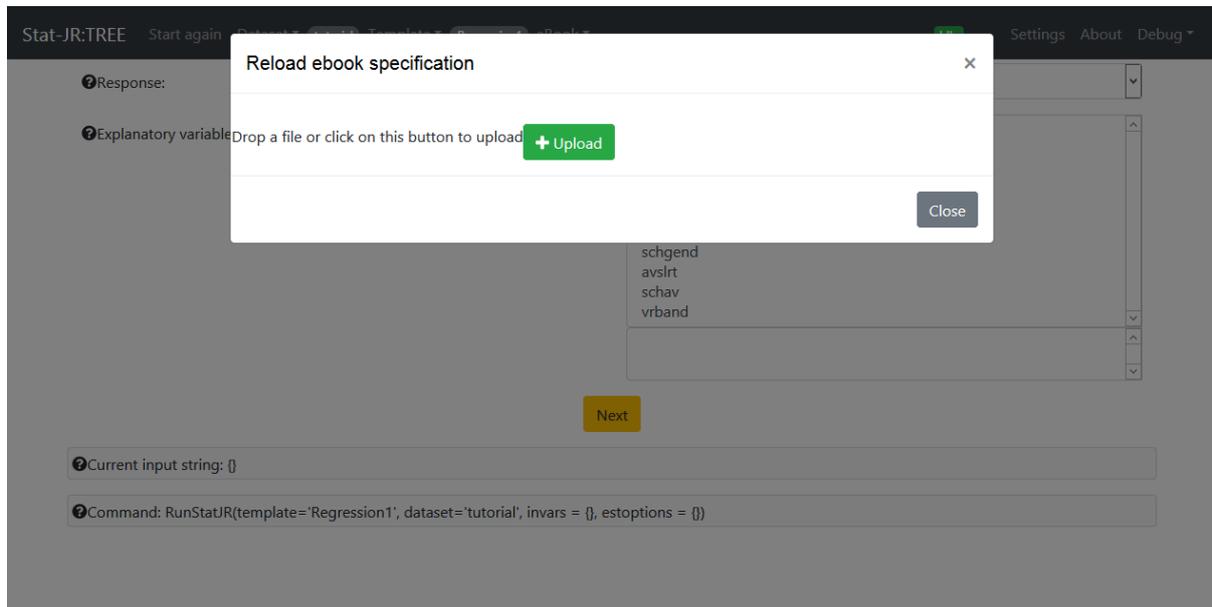
To start using the system click the "TREE" shortcut under "Centre for Multilevel Modelling" within the *Start* menu. Once the application has started the following should appear in a browser window:



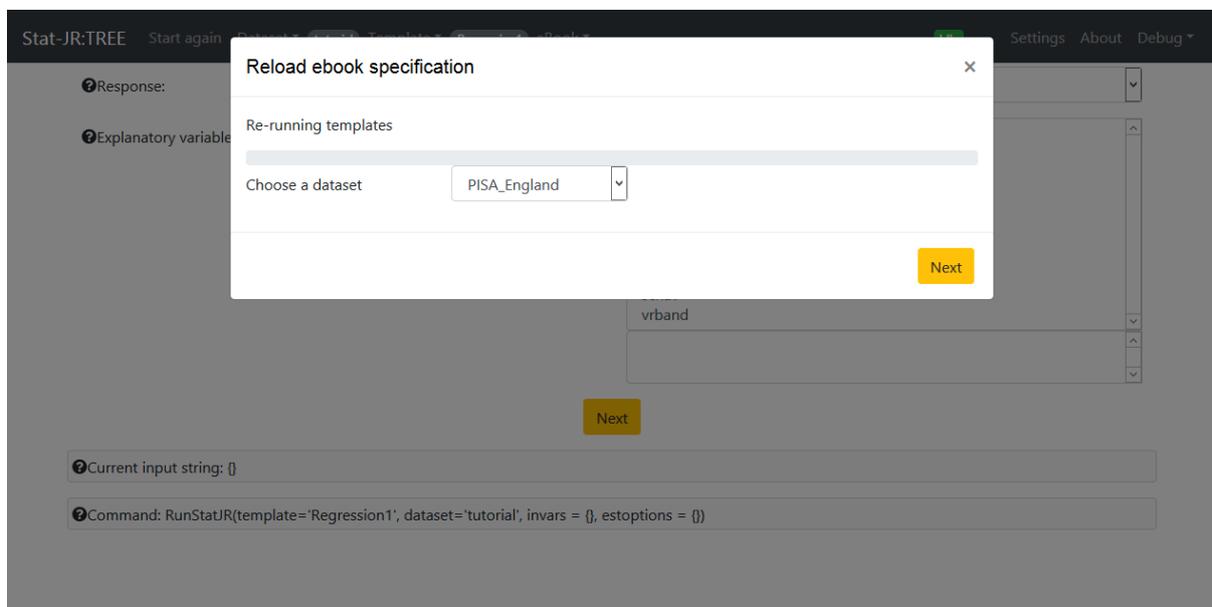
Note that the software has primarily been tested with Chrome and Firefox, so if it opens in a different browser copy the address and use this to open the page in one of these.

While it is possible to create eBooks from scratch using this system we have provided 26 skeletal versions for the 13 topics to help you get started as quickly as possible. For each topic there is a concepts practical eBook which essentially gives the user SPSS instructions and shows the outputs SPSS produces along with interpretations. There is then in addition a quiz (or questions and solutions) practical eBook which consists of two eBook pages, one that asks questions and the other which gives solutions. We will be working with a concepts practical eBook throughout this document.

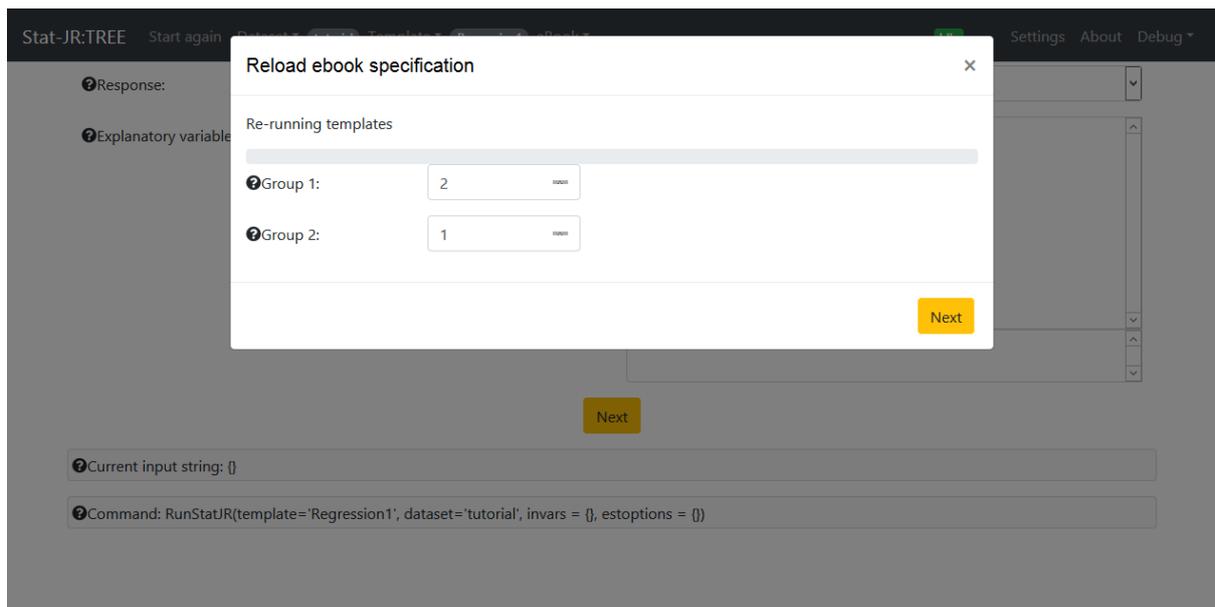
To enable you to become familiar with the system we will start by producing a simple PDF to go with the specific eBook. To do this click the “eBook” menu at the top of the screen, followed by “Load”. You should now see the following screen:



Next click the “+Upload” button, navigate to the location where you put the provided eBooks. We will use *SPSS -Independent Samples t test - Practical.zip* here, so select it and click “Open”. The system will then load and begin to run the eBook. When it encounters an input that has not been specified it will prompt the user to provide an answer specific to this instance of running the eBook. For the provided eBook the first of these will be ask for the data to be used. Select “PISA_England” from the drop-down and click “Next”.

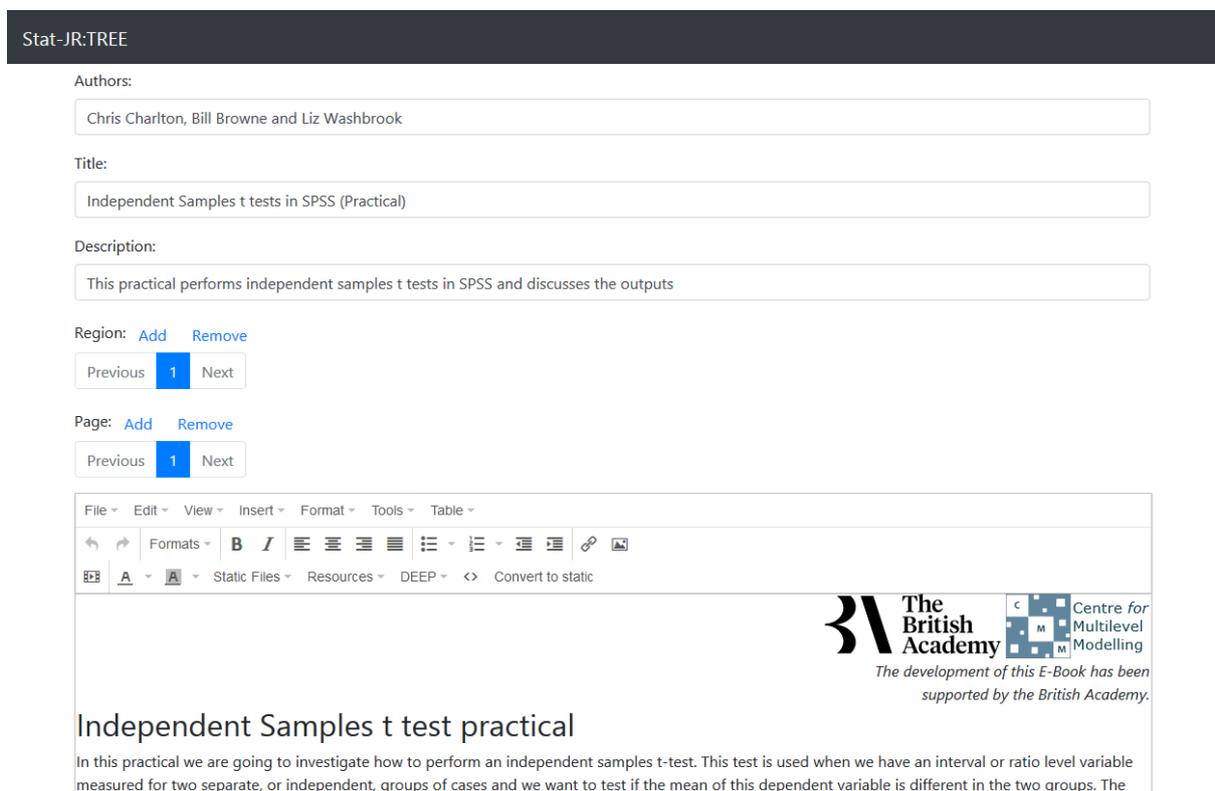


The rest of the questions will ask for the Test Variable and Grouping Variable, so Select “SCISCORE” and “GENDER” respectively clicking “Next” after each input:



Finally choose (type in) “2” and “1” to represent the two groups as shown above and click “Next” again.

Now that all the inputs have been specified the system can finishing running the commands referenced by the eBook, and it will proceed to do so. When this has finished the screen will switch to an eBook editing screen containing the content from the loaded eBook, along with outputs created from running the referenced commands with the chosen data and variable.



This eBook contains the concepts practical that gives instructions and interprets the outputs. At this stage we are simply replicating the provided PDF for this eBook, so to continue scroll to the bottom of the page and click the “Download as eBook” button.

The second SPSS output table contains details of the test itself and can be seen below:

Independent Samples Test									
Levene's Test for Equality of Variances				t-Test for Equality of Means					
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
							Lower	Upper	

In conclusion, we could report this to a reader as follows: Mean test scores were higher among **GENDER = Male** (N=2719, M=527.5540, SD=104.60755) than **GENDER = Female** (N=2475, M=519.2443, SD=100.14185). Levene's test rejected the null hypothesis of equal variances between the two groups (F=8.483, $p < .001$) so an adjusted version of the independent samples t-test that relaxes this assumption was chosen. The difference in means (difference = 8.30970) was statistically significant, $t(5178.827) = 2.924$, $p = .003$.

123 WORDS POWERED BY TINYMCE

Download as eBook

Return to template running environment

You will then be asked for a file name to use for the new eBook. Type in "Ttest1.zip".

Finally click the "Save changes" button. At this stage, depending on your browser settings, the file will either be downloaded to your default download location or you will be asked for a location to save the file. Make a note of this as it will be required for the next stage.

Reading and printing the eBook

We will now load this up in the eBook version of Stat-JR so return to the "Centre for Multilevel Modelling" group within the *Start* menu and select the "DEEP" module. This module provides a read-only reading interface for eBooks. As well as static text this allows reading of more dynamic eBooks with user inputs, however we will not use this functionality in this document.

After a while the following will appear:

Stat-JR:DEEP Import
About

Your E-Books:

About:

Author

Created at

Description

Continue reading:

Start

OR

Start a new reading:

New reading process name:

Brief description:

Start reading

Again, if this opens in a browser other than Chrome or Firefox copy the address and open it in one of these.

To load the eBook that we have saved earlier click the *“Import”* button at the top of the screen.

Stat-JR:DEEP Import
About

Import E-Book
×

+ Select an E-Book file

or Find E-Books on [my](#) [experiment](#)

Your E-Books:

About:

Author

Created at

Description

Continue reading:

Start

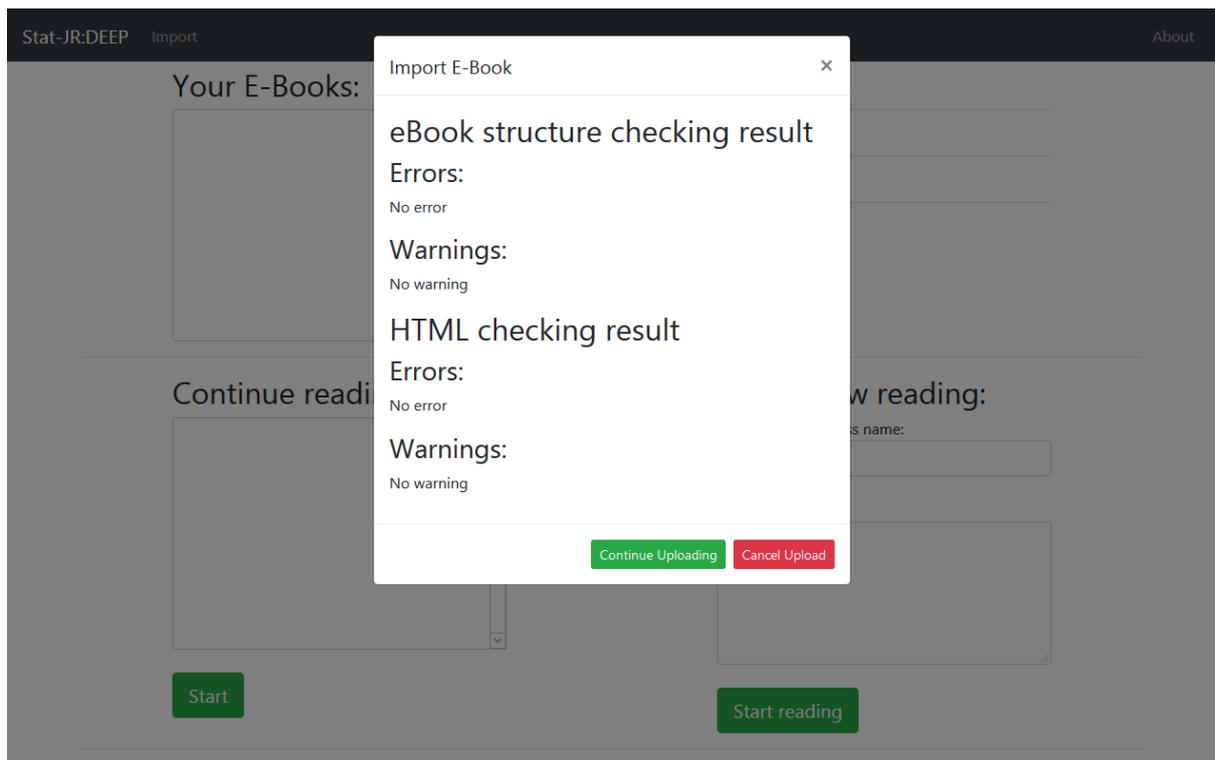
Start a new reading:

New reading process name:

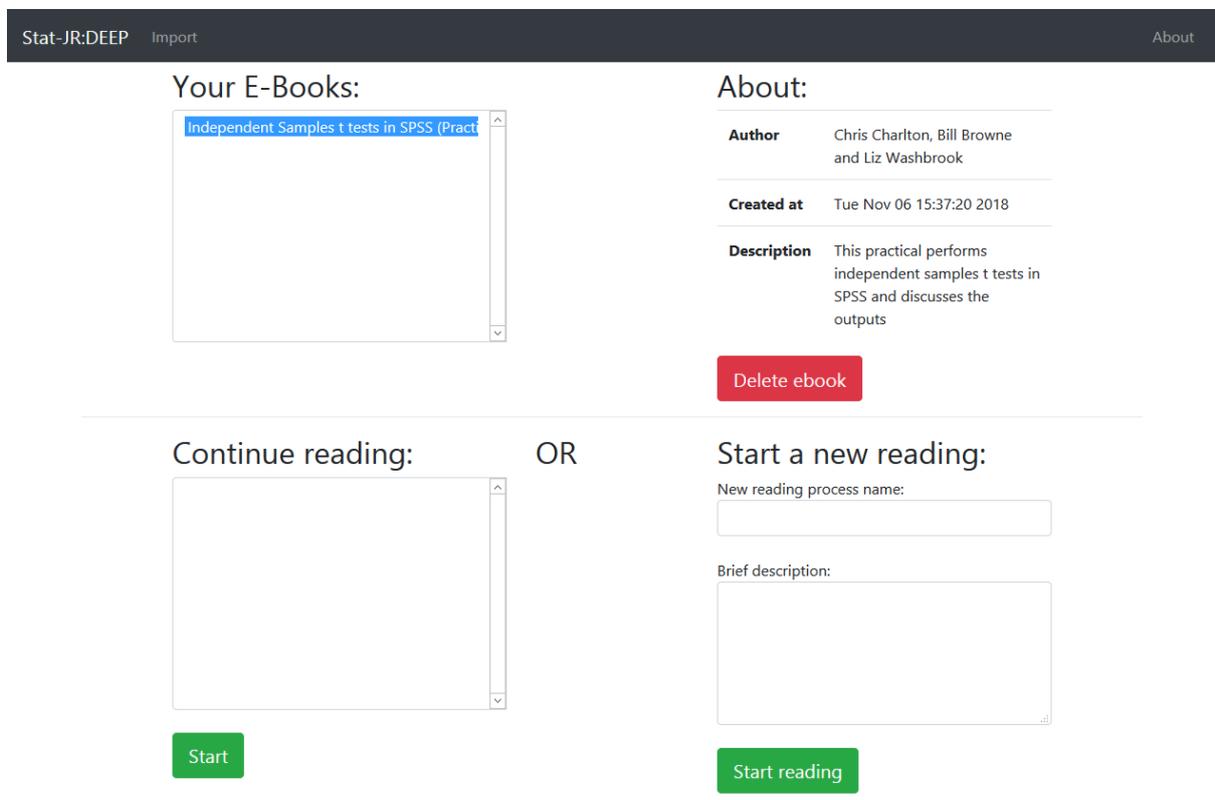
Brief description:

Start reading

Next, click *“Select an E-Book file”* and navigate to the location that you saved the eBook in the previous section, click on the file *Ttest1.zip* and click *“Open”*. Once you have done this the following screen will appear:



Click on “Continue Uploading”, followed by “Continue”. Once this has completed the title of the eBook will appear in the list of available eBooks. Clicking on this will provide additional information regarding the eBook content.



After selecting the eBook type a name into the “New reading process name:” box and click “Start reading”. The purpose of reading processes is to enable multiple instances of an eBook to be run

independently at the same time, and this is therefore most useful if the eBooks in question contain user inputs that might result in the need for multiple different outputs. In our case the output will always be the same, so this is not required. The name and description chosen is only to aid the memory of the user, so what you type here does not matter.

After clicking the “*Start reading*” button the system will switch to the eBook and begin to re-run the referenced commands. Once this has completed we will eventually see the following:

Stat-JR:DEEP Upload Resources About Debug

Finished

« 1 » Go to page

Independent Samples t test practical

The British Academy Centre for Multilevel Modelling
The development of this E-Book has been supported by the British Academy.

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

In this case the groups are identified by specific group codes although it is possible to also use a second continuous variable for groups and specify ranges of this variable that correspond to groups.

Here we are considering the variable, **SCISCORE**.

As a first step we should generally test for the normality of the variable **SCISCORE** in each of the two groups that are indicated by **GENDER**. We first test for normality in each group.

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **Science test score[SCISCORE]** variables into the **Dependent List:** box.
- Copy the **Student gender[GENDER]** variable into the **Factor List:** box.
- Click on the **Plots...** button.

You can now create a PDF file by printing the page to a PDF printer, this is done by right clicking in the window and selecting print ... and then changing destination to *Save as PDF* as shown below:

Print
Total: 4 pages
Save Cancel

Destination: Save as PDF
Change...

Pages: All (selected), e.g. 1-5, 8, 11-13

Margins: Default

Options: Headers and footers, Background graphics

Independent Samples t tests in SPSS (Practical)

The British Academy Centre for Multilevel Modelling

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

In this case the groups are identified by specific group codes although it is possible to also use a second continuous variable for groups and specify ranges of this variable that correspond to groups.

Here we are considering the variable, **SCISCORE**.

As a first step we should generally test for the normality of the variable **SCISCORE** in each of the two groups that are indicated by **GENDER**. We first test for normality in each group.

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **Science test score[SCISCORE]** variables into the **Dependent List:** box.
- Copy the **Student gender[GENDER]** variable into the **Factor List:** box.
- Click on the **Plots...** button.

We will first look at a histogram of the **SCISCORE** variable, for the group defined by **GENDER = Male**. This can be found in amongst the set of output objects and looks as follows:

Histogram for GENDER = Male

Frequency

Science test score

Identify for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case 527.540. This distribution appears to be reasonably symmetric.

Next we will look at a histogram of the **SCISCORE** variable, for the group defined by **GENDER = Female**. This can be found in amongst the set of output objects and looks as follows:

• Select **Descriptive Statistics** from the **Analyze** menu.

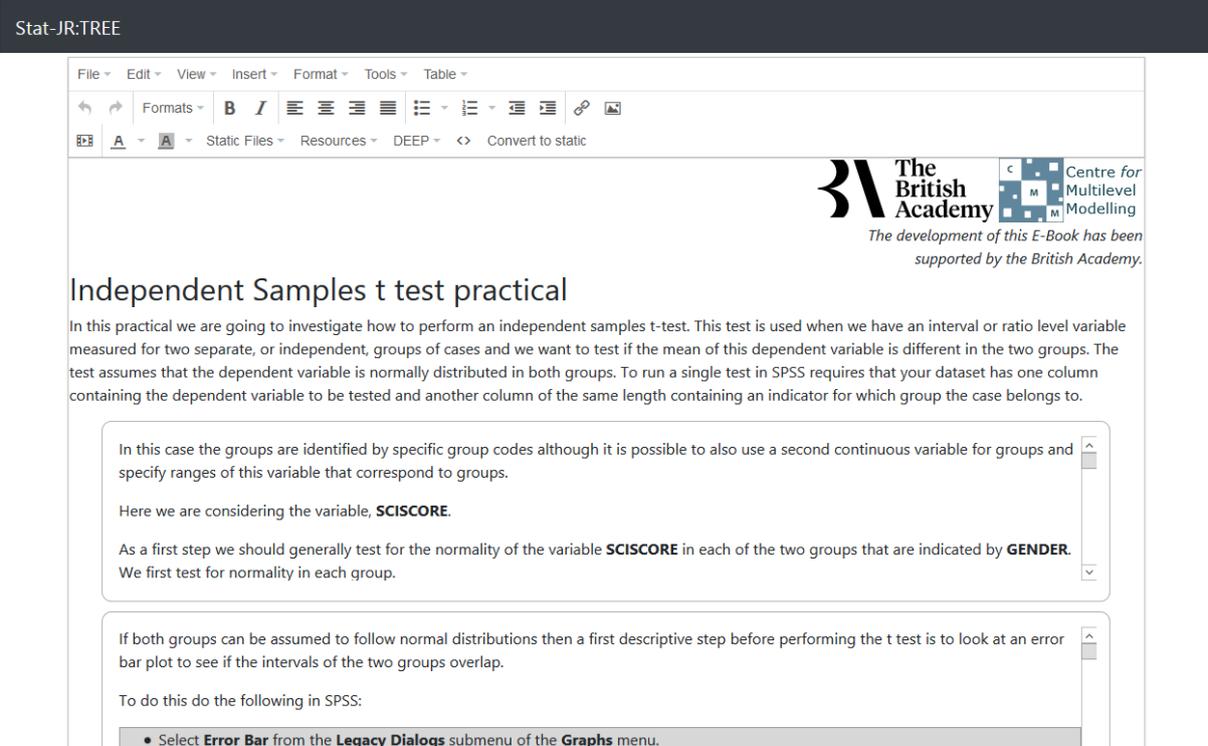
Depending on your existing printer settings you may wish to ensure that headers/footers are turned off and the paper size is appropriate. If you now open the generated PDF file it should match the file that was originally provided. You will see the file is 4 pages long.

Creating this document took only a few button presses but you may want to customise the document in several ways as we'll describe below. One thing that is probably likely is you will also want to run the equivalent Quiz eBook with a different dataset and/or inputs so that you have a different example for the concepts practical and for the quiz practical.

Customisation

1. Exploring the example

To start with we will repeat the operations performed earlier in TREE so see if you can remember what you did and look back if you are unsure but stop BEFORE pressing the *Download as eBook* button. The screen should look as follows:



The screenshot shows a web browser window with the address bar displaying 'Stat-JR:TREE'. The browser's menu bar includes 'File', 'Edit', 'View', 'Insert', 'Format', 'Tools', and 'Table'. The toolbar contains various icons for navigation and editing. The page content includes the logo for 'The British Academy' and the 'Centre for Multilevel Modelling'. The main heading is 'Independent Samples t test practical'. The text explains that the practical will investigate how to perform an independent samples t-test. It notes that the test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The text assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to. There are two scrollable text boxes. The first box contains text about group identification and normality testing. The second box contains text about descriptive statistics and SPSS instructions. A bullet point at the bottom of the second box reads: '• Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.'

Here you will notice some introductory text and then some boxes – for this eBook and page there are 5 boxes and each one does a separate part of the practical. For this practical in order they do the following: in box 1 histograms and normality tests for each variable; in box 2 an error bar plot comparing the two groups; in box 3 a description of the summary statistics for the t test itself; in box 4 a description of the test itself and what it means and in box 5 a description of how one should report the outcome.

To customise the practical we have two routes – firstly we can remove whole chunks of the practical so for example if you don't want to do the histograms and normality tests as you have covered them elsewhere you can click on the first box to highlight it and then press the backspace (delete) button to remove it leaving the screen looking as follows:

File Edit View Insert Format Tools Table

Formats B I [List of icons]

Static Files Resources DEEP Convert to static

The British Academy Centre for Multilevel Modelling
The development of this E-Book has been supported by the British Academy.

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

If both groups can be assumed to follow normal distributions then a first descriptive step before performing the t test is to look at an error bar plot to see if the intervals of the two groups overlap.

To do this do the following in SPSS:

- Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.
- Select **Simple** and **Summaries for groups of cases** as for the boxplot and click on the **Define** button.

We will now do the t test itself.

Below you will see instructions on how to perform the t test in SPSS. If you follow the instructions you will see the two tabular outputs that are embedded in the explanations below.

- Select **Compare Means** from the **Analyze** menu.

Here we now only have 4 boxes and if we were to download this eBook the histograms and normality test section would not appear in the PDF document constructed via DEEP.

We can also add text to the document so for example we might at the start want to say something about the dataset and here we just type some extra words:

File Edit View Insert Format Tools Table

Formats B I [List of icons]

Static Files Resources DEEP Convert to static

The British Academy Centre for Multilevel Modelling
The development of this E-Book has been supported by the British Academy.

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

This is just a test to demonstrate adding text!!!!!!!

If both groups can be assumed to follow normal distributions then a first descriptive step before performing the t test is to look at an error bar plot to see if the intervals of the two groups overlap.

To do this do the following in SPSS:

- Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.
- Select **Simple** and **Summaries for groups of cases** as for the boxplot and click on the **Define** button.

We will now do the t test itself.

Below you will see instructions on how to perform the t test in SPSS. If you follow the instructions you will see the two tabular outputs that are embedded in the explanations below.

Here you will see I have added 1 sentence into the document outside the boxes. A word of warning here is that if you do this inside a box via this route then this will not work as DEEP will recreate the

boxes from scratch. The way around this is to click on the “Convert to static” button at the top of the output and this will simply convert the whole page to a static text object thus:

Stat-JR:TREE

The screenshot shows a web editor interface with a menu bar (File, Edit, View, Insert, Format, Tools, Table) and a toolbar. The document content includes:

- Header: The British Academy Centre for Multilevel Modelling. The development of this E-Book has been supported by the British Academy.
- Section: Independent Samples t test practical
- Text: In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.
- Text: This is just a test to demonstrate adding text!!!!!!!!!!
- Text: If both groups can be assumed to follow normal distributions then a first descriptive step before performing the t test is to look at an error bar plot to see if the intervals of the two groups overlap.
- Text: To do this do the following in SPSS:
 - Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.
 - Select **Simple** and **Summaries for groups of cases** as for the boxplot and click on the **Define** button.
 - Transfer the Science test score[SCISCORE] variable to the **Variable** box.
 - Transfer the **Student gender[GENDER]** variable to the **Category Axis** box.
 - Click on the **OK** button.
- Text: The output will look as follow:

Here you’ll see the boxes have vanished and any changes you make to anything on this page will be saved when you *Download* the eBook.

For example the conclusion text at the bottom of the page (scroll down) is a little suboptimal as it refers to Gender = Male and Gender = Female. We can edit this as follows:

Stat-JR:TREE

The screenshot shows the same web editor interface, but with the conclusion text visible:

The column labelled "Sig (2-tailed)" contains a test of the null hypothesis that the means of the **SCISCORE** variable in the two groups are the same. By default, the two-tailed test reported uses a non-directional alternative hypothesis. It gives the probability that the data in the sample came from a population in which the group means are truly equal, when either a positive or a negative difference between sample group means is evidence against that null hypothesis. To conduct a one-tailed test, in which the alternative hypothesis specifies a particular direction to the difference, we would simply halve the p-value provided by SPSS.

We can reject the null if there is sufficient evidence that the mean of Group 1 is either higher or lower than the mean of Group 2. SPSS looks up the t statistic in the appropriate table gives the associated p value associated with the calculated t-statistic and degrees of freedom. In this case SPSS reports it as $p = .003$. Here we see that the p value is less than 0.05 and therefore we can reject the null hypothesis that the two groups have the same means. Finally we can see the 95% confidence interval for the difference which runs from 2.73837 to 13.88104. Another way to decide whether we can reject the null hypothesis is to check whether this interval contains the value 0. Here we see it does not, a finding that would lead us to reject the null hypothesis.

In conclusion, we could report this to a reader as follows: Mean test scores were higher among **Males** (N=2719, M=527.5540, SD=104.60755) than **Females** (N=2475, M=519.2443, SD=100.14185). Levene's test rejected the null hypothesis of equal variances between the two groups ($F=8.483$, $p < .001$) so an adjusted version of the independent samples t-test that relaxes this assumption was chosen. The difference in means (difference = 8.30970) was statistically significant, $t(5178.827) = 2.924$, $p=.003$.

At the bottom of the page, there are two buttons: "Download as ebook" and "Return to template running environment".

Now if you *Download* the eBook and save it as *Ttest2.zip* you can look at it in DEEP. Note you should delete the current eBook in DEEP and then follow the previous instructions. The output will be as shown in the screenshots below:

Independent Samples t tests in SPSS (Practical)

Idle

 « 1 » Go to page

[Independent Samples t test practical](#)


Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

This is just a test to demonstrate adding text!!!!!!!

If both groups can be assumed to follow normal distributions then a first descriptive step before performing the t test is to look at an error bar plot to see if the intervals of the two groups overlap.

To do this do the following in SPSS:

- Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.
- Select **Simple** and **Summaries for groups of cases** as for the boxplot and click on the **Define** button.
- Transfer the Science test score[SCISCORE] variable to the **Variable** box.
- Transfer the **Student gender**[GENDER] variable to the **Category Axis** box.
- Click on the **OK** button.

The output will look as follow:

and scrolling down:

Independent Samples t tests in SPSS (Practical)

Idle

 « 1 » Go to page

[Independent Samples t test practical](#)

reversed the definition of the groups in the initial dialog box, in this case the mean difference would have been negative -8.3097. Next to the mean difference is the standard error of the difference. This here has the value 2.84190 and is calculated via a formula from the standard errors of the means of each group and their respective sample sizes. You will notice that this value is slightly different from the value in the row above for results with Equal variances assumed, which results from the adjustment made for unequal variances. Working back to the start of the row, the column entitled t is the statistic used in the t test and t like F is a standard statistical distribution. The t statistic is calculated simply by dividing the mean difference by its standard error so $8.30970 / 2.84190 = 2.924$. Next to t is a column labelled df which stands for degrees of freedom and is a parameter used to choose the correct t distribution for the sampling distribution of the statistic. When we can assume equal variances then the degrees of freedom equal two less than the number of observations (N - 2, here 5192) as we have used 2 degrees of freedom in estimating 2 means. If we cannot assume equal variances then the value is lower as seen here (5178.827).

The column labelled "Sig (2-tailed)" contains a test of the null hypothesis that the means of the **SCISCORE** variable in the two groups are the same. By default, the two-tailed test reported uses a non-directional alternative hypothesis. It gives the probability that the data in the sample came from a population in which the group means are truly equal, when either a positive or a negative difference between sample group means is evidence against that null hypothesis. To conduct a one-tailed test, in which the alternative hypothesis specifies a particular direction to the difference, we would simply halve the p-value provided by SPSS.

We can reject the null if there is sufficient evidence that the mean of Group 1 is either higher or lower than the mean of Group 2. SPSS looks up the t statistic in the appropriate table gives the associated p value associated with the calculated t-statistic and degrees of freedom. In this case SPSS reports it as $p = .003$. Here we see that the p value is less than 0.05 and therefore we can reject the null hypothesis that the two groups have the same means. Finally we can see the 95% confidence interval for the difference which runs from 2.73837 to 13.88104. Another way to decide whether we can reject the null hypothesis is to check whether this interval contains the value 0. Here we see it does not, a finding that would lead us to reject the null hypothesis.

In conclusion, we could report this to a reader as follows: Mean test scores were higher among **Males** (N=2719, M=527.5540, SD=104.60755) than **Females** (N=2475, M=519.2443, SD=100.14185). Levene's test rejected the null hypothesis of equal variances between the two groups ($F=8.483$, $p < .001$) so an adjusted version of the independent samples t-test that relaxes this assumption was chosen. The difference in means (difference = 8.30970) was statistically significant, $t(5178.827) = 2.924$, $p=.003$.

2. Reproducing the supplied PISA PDF

For each of the 26 generic eBooks that have been written we have constructed pdf files using the PISA data that you have been exploring here but adding in a paragraph of text at the start and end of the practical to put the material into context. To replicate this process is a very similar exercise as was done in customisation part 1. To start with repeat running the eBook in TREE to get to the eBook editor screen as you see at the bottom of page 3 and repeated here below:

Authors:

Chris Charlton, Bill Browne and Liz Washbrook

Title:

Independent Samples t tests in SPSS (Practical)

Description:

This practical performs independent samples t tests in SPSS and discusses the outputs

Region: [Add](#) [Remove](#)Previous **1** NextPage: [Add](#) [Remove](#)Previous **1** Next

File Edit View Insert Format Tools Table

Formats **B** *I* [List Icons] [Link Icon]

Static Files Resources DEEP <> Convert to static

The British Academy Centre for Multilevel Modelling
The development of this E-Book has been supported by the British Academy.

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The

If you have looked at the PDFs for the PISA materials you will note that we use blue font to indicate the contextual information. The font colour can be reached by clicking on the symbol indicated below and then the text can be added thus:

File Edit View Insert Format Tools Table

Formats **B** *I* [List Icons] [Link Icon]

Static Files Resources DEEP <> Convert to static

The British Academy Centre for Multilevel Modelling
The development of this E-Book has been supported by the British Academy.

Independent Samples t test practical

In this practical we are going to investigate how to perform an independent samples t-test. This test is used when we have an interval or ratio level variable measured for two separate, or independent, groups of cases and we want to test if the mean of this dependent variable is different in the two groups. The test assumes that the dependent variable is normally distributed in both groups. To run a single test in SPSS requires that your dataset has one column containing the dependent variable to be tested and another column of the same length containing an indicator for which group the case belongs to.

The example provided here explores the question of whether there are gender differences in the PISA measure of achievement in science. The variable SCISCORE is the dependent variable (individual's test scores) and the group indicator is provided by the GENDER variable (with categories Male and Female).

In this case the groups are identified by specific group codes although it is possible to also use a second continuous variable for groups and specify ranges of this variable that correspond to groups.

Here we are considering the variable, **SCISCORE**.

As a first step we should generally test for the normality of the variable **SCISCORE** in each of the two groups that are indicated by **GENDER**.

Scrolling to the bottom of the practical we have repeated the process to add some blue conclusions text thus:

Independent Samples Test									
Levene's Test for Equality of Variances			t-test for Equality of Means				95% Confidence Interval of the Difference		
F	Sig.		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
<p>In conclusion, we could report this to a reader as follows: Mean test scores were higher among GENDER = Male (N=2719, M=527.5540, SD=104.60755) than GENDER = Female (N=2475, M=519.2443, SD=100.14185). Levene's test rejected the null hypothesis of equal variances between the two groups (F=8.483, $p < .001$) so an adjusted version of the independent samples t-test that relaxes this assumption was chosen. The difference in means (difference = 8.30970) was statistically significant, $t(5178.827) = 2.924$, $p = .003$.</p> <p>According to the PISA test, boys in England were significantly better at science than girls at the age of 15.</p>									

P > SPAN 189 WORDS POWERED BY TINYMCE

[Download as ebook](#)
[Return to template running environment](#)

If you now download this eBook and print the PDF from DEEP you should get an exact copy of the PDF supplied with the software. We do also supply an eBook that already contains this text.

3. Using your own dataset

Now that we have examined recreating and modifying the example materials we next describe how to apply the same eBook to your own data.

Firstly, it is necessary for the system to be able to find your data. To do this you need to save it to a Stata ".dta" file and place it in your Stat-JR personal data directory. This can be changed with the settings in TREE, however the default location is %USERPROFILE%\statjr\datasets. Pasting this address into Windows explorer should take you directly to the folder. Once the data is in the correct location then either restart TREE if you closed it previously or, if you are still in the eBook editor click "*Return to template running environment*" button at the bottom of the screen followed by selecting "*Reload datasets*" from the "*Debug*" menu in the top right-hand corner.

To create an eBook using your new data use the "*Load*" option under the "*eBook*" menu as before to load the initial eBook, but this time select your data from the drop-down and choose an appropriate variable to use as your response.

Having done this the commands will be run with your data and you will be returned to the eBook editing screen. You will notice that some of the text and other outputs will be different to that seen before, to match the new data.

As the eBook contents have now changed you will need to do some editing before saving the final eBook. Start by adding your name to the authors box at the top of the page. The contents of the "*Title*" box is what appears in the eBook selection list within the DEEP interface, so it is also important that you change this to allow easy selection of your new eBook later.

Having changed the descriptive information regarding the template we can now move on to editing the contents. You will notice that the page consists of text interspersed with outputs within boxes. As mentioned in the last section these boxes contain dynamic text generated from the SPSS outputs and are not editable by default. The main text can be altered using a standard word-processing type interface, and extra SPSS outputs can be added if necessary from the Resources menu within the editor.

If you only change the static text then when you run the eBook in future all the outputs will be regenerated using the current version of your data. This allows you to make modifications to your data in the future and have the eBook automatically update to reflect this. There may however be cases where you wish to make changes to the dynamic text, in which case you can do so with the “Convert to static” button at the top of the editing box as illustrated earlier. After doing this the outputs will no longer reflect any changes in the underlying data. Another change is if you make all 3 pages static then when you load the eBook in DEEP SPSS will no longer be run in the background as all output values are already known.

When you are happy with the changes to your eBook save it as before (using a different name). You can then load this into the DEEP environment in the same way as in the first section and create a PDF.

Appendix – Background to the PISA datafile

The datafiles supplied with the learning materials and used in this document is an extract from the Programme for International Student Assessment (PISA), a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. In 2015 over half a million students, representing 28 million 15-year-olds in 72 countries and economies, took the internationally agreed two-hour test. Students were assessed in science, mathematics, reading, collaborative problem solving and financial literacy. Background information was collected in questionnaires administered to students and school principals and, in some countries, to parents.

The accompanying datafiles contain selected variables from the student questionnaire and derived test scores, from the samples from England (*PISA_England.dta*; used in the example materials provided) and from (South) Korea (*PISA_Korea.dta*; a supplemental file to allow for generation of a parallel set of learning materials by the user). Access to the underlying data are unrestricted, with documentation provided at <http://www.oecd.org/pisa/data/2015database/> and the link to the full 72-country Student questionnaire data file at http://vs-web-fs-1.oecd.org/pisa/PUF_SPSS_COMBINED_CMB_STU_QQQ.zip.

The SPSS syntax file *PISA_extract.sps* (available from <http://www.bristol.ac.uk/cmm/software/statjr/downloads/>) creates the derived files from the input file *CY6_MS_CMB_STU_QQQ.sav* (contained in the zip file downloaded from the location given above). The variable *Region* identifies the country or region to which a record belongs, and can therefore be used to generate an equivalent derived datafile for any country selected by the user (by editing the corresponding line in the SPSS syntax file). For example, retaining records for which *Region* = 82611 selects the England sub-sample and those for which *Region* = 41000 selects the Korea sub-sample. The syntax file selects, renames, re-labels and recodes variables to produce the output datafiles for use with the learning materials. The output datafile is automatically saved in two versions: in SPSS format (with a .sav extension) which is the version that should be circulated to learners along with the pdfs of the learning materials; and in Stata format (with a .dta extension) which is used for generating the learning materials in the Stat-JR software (SPSS .sav datafiles cannot be read directly into the Stat-JR software).

It should be noted that certain features of SPSS datafiles do not transfer to Stata datafiles, so original SPSS files may require some modification before they are appropriate to use with the generated learning materials. Specifically, SPSS allows a range of codes to designate missing values and these all must be converted to “system missing” (denoted by .) in files that accompany the learning materials (using the SPSS syntax `RECODE ALL (MISSING=SYSMIS)` which for example is used in the file *PISA_extract.sps*). In addition, variable names must not exceed 32 characters and variable labels must not exceed 80 characters.

The table below provides details of the variables supplied in the accompanying datasets. References to further information are provided in the table (e.g. Ch. 16, pp.318-319), and refer to the PISA 2015 Technical Report (available at: <http://www.oecd.org/pisa/data/2015-technical-report/>). Three features of the PISA-supplied variables require explanation here.

Item response theory (IRT) derived scales. A number of scales provided in the dataset are PISA-derived variables based on IRT techniques (e.g. EMOSUPS, JOYSCIE). The resulting scales are weighted sums of the underlying items and are constructed to have a mean of 0 and a standard deviation of 1 across OECD countries (with equally weighted countries). The weights applied to items depend on average responses (item “difficulty”) and correlation with the underlying latent construct

(item “discrimination”). Full details of the methodology and weights used are provided in Chapter 16 of the PISA 2015 Technical Report, but perhaps the key to interpretation is to bear in mind that, for all IRT-scaled predictor variables **“the average OECD student would have an index value of zero and about two-thirds of the OECD student population would be between the values of -1 and 1”** (p.293).

Plausible values. PISA achievement tests rely on a large bank of question items, with individual students answering only partial subsets of all available items. An individual’s set of responses is then used to estimate their underlying latent achievement level, a process that involves some degree of uncertainty and error. To reflect this measurement uncertainty, PISA provides not just one estimate of a student’s “true” test score, but ten different “plausible values”, or equally reasonable estimates of an individual’s ability in, say, science. In advanced statistical analyses, the full set of plausible values can be employed to improve accuracy of inference. As the purpose of the data extraction here is to facilitate the teaching of introductory statistical concepts, we provide only the first plausible value for each test score, treating it as an observed value rather than an estimate constructed from a more complicated statistical model. See Chapter 9 of the PISA 2015 Technical Report for further details.

Achievement scales. The overall PISA science score was scaled to have a mean of 500 and a standard deviation of 100 across the population of 15-year olds in the OECD in 2006 (*PISA 2015 Results Volume I*, OECD, p.58). Other dimensions of achievement are scaled in a similar way. PISA guidance is that 30 PISA points are approximately equivalent to the progress made in a year of schooling, on average across the OECD (*ibid*, p.65). A score of 410 is required to reach “Level 2 proficiency” which is considered “the baseline level of science proficiency that is required to engage in science-related issues as a critical and informed citizen” (*ibid*, p.34 and p.68).

Variable table

Variable name	Original PISA variable name	Description	Coding
GENDER	ST004D01T	Student gender	(1) Female; (2) Male
PARINTSCH	ST123Q01NA	My parents are interested in my school activities	(1) Strongly disagree; (2) Disagree; (3) Agree; (4) Strongly agree
PARSUPED	ST123Q02NA	My parents support my educational efforts and achievements	As above
PARSUPDIF	ST123Q03NA	My parents support me when I am facing difficulties at school	As above
PARCONF	ST123Q04NA	My parents encourage me to be confident	As above
EMOSUPS	EMOSUPS	Parental emotional support score	IRT derived scale from items: PARINTSCH; PARSUPED; PARSUPDIF; PARCONF [Ch. 16, pp. 317-318]
INFGGAS	ST092Q01TA	How informed are you about the following environmental issues... The increase of greenhouse gases in the atmosphere	(1) I have never heard of this; (2) I have heard about this but I would not be able to explain what it is really about; (3) I know something about this and could explain the general issue; (4) I am familiar with this and I would be able to explain this well
INFGMO	ST092Q02TA	How informed are you about the following environmental issues... The use of genetically modified organisms (GMOs)	As above
INFNUCL	ST092Q04TA	How informed are you about the following environmental issues... Nuclear waste	As above
INFDEFOR	ST092Q05TA	How informed are you about the following environmental issues... The consequences of clearing forests for other land use	As above
INFAIR	ST092Q06NA	How informed are you about the following environmental issues... Air pollution	As above
INFEXT	ST092Q08NA	How informed are you about the following environmental issues... Extinction of plants and animals	As above

Variable name	Original PISA variable name	Description	Coding
INFWAT	ST092Q09NA	How informed are you about the following environmental issues... Water shortage	As above
OPTAIR	ST093Q01TA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? Air pollution	(1) Get worse; (2) Stay about the same; (3) Improve
OPTEXT	ST093Q03TA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? Extinction of plants and animals	As above
OPTDEFOR	ST093Q04TA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? The consequences of clearing forests for other land use	As above
OPTWAT	ST093Q05TA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? Water shortage	As above
OPTNUCL	ST093Q06TA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? Nuclear waste	As above
OPTGGAS	ST093Q07NA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? The increase of greenhouse gases in the atmosphere	As above

Variable name	Original PISA variable name	Description	Coding
OPTGMO	ST093Q08NA	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years? The use of genetically modified organisms (GMOs)	As above
FUN_4	ST094Q01NA	I generally have fun when I am learning broad science topics	(1) Strongly disagree; (2) Disagree; (3) Agree; (4) Strongly agree
LIKEREAD_4	ST094Q02NA	I like reading about broad sciences	As above
HAPPY_4	ST094Q03NA	I am happy working on broad science topics	As above
ENJKNOW_4	ST094Q04NA	I enjoy acquiring new knowledge in broad science topics	As above
INTEREST_4	ST094Q05NA	I am interested in learning about broad science	As above
JOYSCIE	JOYSCIE	Enjoyment of science score	IRT derived scale from items: FUN_4, LIKEREAD_4, HAPPY_4, ENJKNOW_4, INTEREST_4. [Ch. 16, pp. 310-312]
INT_BIO	ST095Q04NA	To what extent are you interested in... Biosphere (e.g. ecosystem services, sustainability)	(1) Not interested; (2) Hardly interested; (3) Interested; (4) Highly interested
INT_FORCES	ST095Q07NA	To what extent are you interested in... Motion and forces (e.g. velocity, friction, magnetic and gravitational forces)	As above
INT_ENERGY	ST095Q08NA	To what extent are you interested in... Energy and its transformation (e.g. conservation, chemical reactions)	As above
INT_UNIV	ST095Q13NA	To what extent are you interested in... The Universe and its history	As above
INT_DISEASE	ST095Q15NA	To what extent are you interested in... How science can help us prevent disease	As above
INTSCI	INTBRSCI	Interest in science topics score	IRT derived scale from items: INT_BIO, INT_FORCES, INT_ENERGY, INT_UNIV, INT_DISEASE. [Ch. 16, pp. 310-312]
USEFUL_4	ST113Q01TA	Making an effort in my school science subject(s) is worth it because this will help me in the work I want to do later on	(1) Strongly disagree; (2) Disagree; (3) Agree; (4) Strongly agree

Variable name	Original PISA variable name	Description	Coding
NEEDED_4	ST113Q02TA	What I learn in my school science subject(s) is important for me because I need this for what I want to do later on	As above
CAREER_4	ST113Q03TA	Studying my school science subject(s) is worthwhile for me because what I learn will improve my career prospects	As above
JOB_4	ST113Q04TA	Many things I learn in my school science subject(s) will help me to get a job	As above
INSMOVSCI	INSTSCIE	Science instrumental motivation score	IRT derived scale from items: USEFUL_4, NEEDED_4, CAREER_4, JOB_4. [Ch. 16, pp. 313 & 315]
SCIEEFF	SCIEEFF	Science self-efficacy score	IRT derived scale from 8 items: How easy do you think it would be for you to perform the following tasks on your own? Recognise the science question that underlies a newspaper report on a health issue. Explain why earthquakes occur more frequently in some areas than in others. Describe the role of antibiotics in the treatment of disease. Identify the science question associated with the disposal of garbage. Predict how changes to an environment will affect the survival of certain species. Interpret the scientific information provided on the labelling of food items. Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars. Identify the better of two explanations for the formation of acid rain. [Ch. 16, pp. 318-319]
SCIEACT	SCIEACT	Science activities index	IRT derived scale from 9 items: How often do you do these things? Watch TV programmes about science Borrow or buy books on science topics Visit web sites about science topics

Variable name	Original PISA variable name	Description	Coding
			<p>Read science magazines or science articles in newspapers</p> <p>Attend a science club</p> <p>Simulate natural phenomena in computer programs/virtual labs</p> <p>Simulate technical processes in computer programs/virtual labs</p> <p>Visit web sites of ecology organisations</p> <p>Follow news of science, environmental, or ecology organizations via blogs and microblogging. [Ch. 16, pp. 318-319]</p>
IMMIG	IMMIG	Immigration status	(1) Native; (2) Second-Generation; (3) First-Generation
CULTPOSS	CULTPOSS	Home cultural possessions score	IRT derived scale from 5 items recording presence in household of: <ul style="list-style-type: none"> Classical literature Books of poetry Works of art Books on art, music or design Musical instruments [Ch. 16, pp. 300-305]
HEDRES	HEDRES	Home educational resources score	IRT derived scale from 7 items recording presence in household of: <ul style="list-style-type: none"> A desk to study at A quiet place to study A computer you can use for schoolwork Educational software Books to help you with school work Technical reference books A dictionary [Ch. 16, pp. 300-305]
WEALTH	WEALTH	Family wealth score	IRT derived scale from 12 items recording presence in household of: <ul style="list-style-type: none"> A room of your own A link to the internet Televisions Cars Rooms with bath or shower Smartphones

Variable name	Original PISA variable name	Description	Coding
			Computers Tablet computers Ebook readers 3 country-specific wealth indicators (e.g. for England: a premium TV package, a HD TV, and a tablet computer) [Ch. 16, pp. 300-305]
HOMEPOS	HOMEPOS	Home possessions score	IRT derived scale from the 24 constituent items of CULTPOSS, HEDRES and WEALTH, plus item on number of books in home [Ch. 16, pp. 300-305]
ESCS	ESCS	Economic, social and cultural status index	Composite indicator based on three PISA-derived variables: HOMEPOS, HISEI (highest value for a parent on the International Socio-Economic Index of Occupational Status) and PARED (highest number of years a parent spent in education) [Ch. 16, pp. 339-342]
MATHSCORE	PV1MATH	Math test score	Test score scaled to $M=500$, $SD=100$ across the OECD
READSCORE	PV1READ	Reading test score	As above
SCISCORE	PV1SCIE	Science test score	As above
SCI_PHYS	PV1SSPH	Physical systems sub-score	As above
SCI_LIVING	PV1SSLI	Living systems sub-score	As above
SCI_EARTH	PV1SSES	Earth & space systems sub-score	As above
PAREDU	HISCED	Highest qualification of parent	Recoded from HISCED to (approx) equivalents used in England. HISCED categorises qualifications according to the International Standard Classification of Education (ISCED): None, ISCED1, 2, 3C, 3B = (1) Low: GCSE or equiv ISCED 3A, 4, 5B = (2) Medium: A-level or equiv ISCED 5A, 6 = (3) High: University degree